Original Paper

Phonetica 2001;58:129–162

Received: October 28, 1999
Accepted after revision: October 9, 2000

Effects of Morpheme Boundaries on Intergestural Timing: Evidence from Korean

Taehong Cho

Department of Linguistics, University of California, Los Angeles, Calif., USA

Abstract

This paper examines the effects of morpheme boundaries on intergestural timing, and demonstrates that low-level phonetic realization is influenced by morphological structure, i.e. compounding and affixation. It reports two experiments, one using electromagnetic midsagittal articulography (EMA) and one electropalatography (EPG), examining Korean data. The results of the EMA study show that intergestural timing is less variable for adjacent gestures across the word boundary inside a lexicalized compound than inside a nonlexicalized compound, and inside a monomorphemic word than across a morpheme boundary. The EPG study (which examined the timing in palatalization of a coronal) shows that both [ti] and [ni] have more variability in gestural timing when heteromorphemic than when tautomorphemic. Furthermore, the phonetic details of gestural overlap shed light on the asymmetry on palatalization between tautomorphemic and heteromorphemic gestural sequences (e.g. *ni* vs. *n-i*), presumably driven by paradigmatic contrast and preference of overlap. In short, what emerges from two experiments is that gestures are coordinated more stably within a single lexical item (a morpheme or a lexicalized compound) than across a boundary between lexical items. In accounting for the stability of intergestural timing within a lexical entry, several hypotheses were discussed including the Phase Window, Bonding Strength, Phonological Timing and Extended Phase Window model newly proposed here. The implication is that the morphological structure may be encoded in the phonetic realization, as is the case with other linguistic structure (e.g. prosodic structure).

Introduction

One of the fundamental goals in the field of speech production is the understanding of which aspects of linguistic structure influence low-level phonetic implementation. Recent evidence has demonstrated that phonetic realization of segments or features varies with prosodic structure. Specifically, many researchers have explored spatiotemporal acoustic or articulatory properties at different prosodic boundaries (phrase, word and syllable), and have reached a general consensus that there are no

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/01/0583–0129$17.50/0
Accessible online at:
www.karger.com/journals/pho

Taehong Cho
Department of Linguistics, University of California
Los Angeles, 405 Hilgard Avenue
Los Angeles, CA 90095 (USA)
Tel. +1 310 206-2808, Fax +1 310 206-5743
E-Mail taehong@humnet.ucla.edu

invariant phonetic values associated with a single feature or a segment, and that low-level phonetic phenomena cannot be understood apart from high-level prosodic structure [e.g. Edwards et al., 1991; Pierrehumbert and Talkin, 1992; Jun, 1995; Fougeron and Keating, 1997; Byrd and Saltzman, 1998; Cho, 1999; Cho and Keating, 1999, to appear; Keating et al., 1999, to appear; Byrd, 2000; Byrd et al., 2000]. However, while there is a relatively large body of literature about effects of prosodic structure (e.g. syllable, word and phrase boundaries) on phonetic realizations, few studies have addressed effects of morphological structure. This study explores the possibility that morphological structure does indeed influence phonetic realization in at least some cases, specifically compounding and affixation in Korean.

A good place to look for such effects is in the domain of articulatory overlap: effects can be precisely measured, and there is a body of theoretical literature that can be drawn on in interpreting experimental findings. One such theoretical framework is Articulatory Phonology [Browman and Goldstein, 1986, 1989, 1990, 1992a, b, 1995, 1998], in which spatiotemporal characteristics of articulatory movements are linguistically significant in various ways. In Articulatory Phonology, the articulatory gestures are linguistically meaningful units: the primitives of phonological representation are gestures, not phonemes. In this framework, a variety of linguistic messages (e.g. lexical distinctions, lexical stress, phrasing, etc.) are conveyed through two mechanisms: (1) what kind of gestures are involved and (2) how they are spatiotemporally coordinated. Lexical representations therefore contain not only information about articulators and their movements, but also a timing specification for those gestures, constituting a lexical entry, which is stored in a 'gestural score' [Browman and Goldstein, 1990, 1998]. A key assumption in Articulatory Phonology relevant to the current study is that information about temporal coordination for a lexical item is stored in the lexicon, and that the phonological structure of a lexical item may be viewed as a 'constellation' of gestures; that is, a stable organization among gestures [e.g. Browman and Goldstein, 1986].

Evidence for the stable organization among gestures within a lexical item is found in a variety of studies which show that intergestural timing is more variable at a level above the word than word-internally [e.g. McLean, 1973; Hardcastle, 1985; Holst and Nolan, 1995; Byrd, 1998; Byrd et al., 2000]. Byrd et al. [2000] reported that the labial and lingual gestures for an /nm/ or an /mn/ sequence were phased with each other with less overlap and greater variability when the gestural sequence spanned a major prosodic boundary (e.g. a phrase boundary) as opposed to a lower level boundary (e.g. a syllable boundary within a word). Lesser variability was also found for intragestural timing of lip opening and closing gestures associated with /ma#m/ when the intervening prosodic boundaries (#) were lower, i.e. the word-internal sequence has lesser variability than sequences spanning a higher boundary [Byrd and Saltzman, 1998]. Furthermore, there is a subtle but significant word-level effect on intergestural timing. For example, Byrd [1994, p. 76], in an electropalatographic study of timing patterns in adjacent consonants in English, found that 'coordination of articulatory movements for sequential consonants is less variable if the consonants are tautosyllabic than if they are heterosyllabic'. However, Byrd's [1994] examples for tautosyllabic consonants in fact occur within a word, and the examples for heterosyllabic consonants occur across a word boundary. Thus, Byrd's findings may be interpreted as either a word-level or syllable-level effect.

The current study examines the effect of morpheme and word boundaries on intergestural timing in Korean. The central hypothesis of this paper is that intergestural

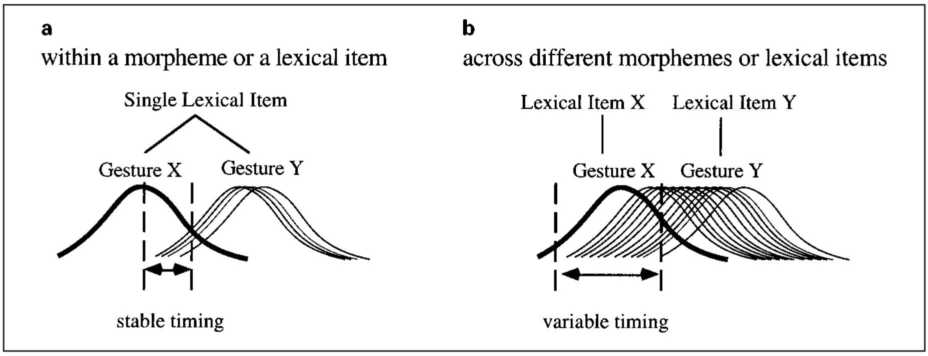

Fig. 1. Different width phase windows for coordinating gestures within a single word (**a**) or in separate words or morphemes (**b**).

timing is sensitive to the morphological structure, such as affixal morpheme or compound boundaries. As a starting point, one prediction can be made, based on Browman and Goldstein [1986]'s view of a lexical item as a stable constellation of gestures, i.e. timing relations are more stable between gestures belonging to a lexical item, as in figure 1a, than between gestures across different lexical items, as in figure 1b.

At this point, it should be clarified what constitutes a lexical item. It will be assumed throughout the paper that each productively used morpheme constitutes a separate lexical entry [cf. Booij, 1985] and therefore, productively affixed forms consist of multiple lexical entries. Thus, if we follow the idea of stable timing within a lexical item in Articulatory Phonology, intergestural timing is expected to be more stable within a morpheme than across morphemes. Furthermore, we can expect intergestural timing to be more stable within a lexicalized compound than across a word boundary in a nonlexicalized compound (or a noun phrase).

A competing view, however, is that there should be no morpheme boundary effects on the phonetic realization of gestures inside a phonological word. One may argue that the phonological form which is the output of phonology is fed into the phonetic component of the grammar without reference to morphological information. One such view can be found in Chomsky and Halle [1968], in which the output of phonology has no information concerning word-internal morphological structures, i.e. the boundaries are erased at the end of each transformational cycle, and the speech signal is generated by applying universal phonetic implementation rules to the output of phonology, i.e. it assumes the relative timing of phonemes to be handled automatically by the phonetic module with no effects of morpheme boundaries. Similarly, the Bracket Erasure Convention in Lexical Phonology [e.g. Kiparsky, 1982; Mohanan, 1982] predicts that morphological structure is invisible to phonetics. Most recently, Levelt et al. [1999] proposed a speech production model in which 'phonetic encoding' does not have access to morphological information. In their model, a string of morphemes (the result of 'morphological encoding') is fed into 'phonological encoding' in which syllabification takes place. At the stage of phonetic encoding, the same 'ready-made' gestural syllabic score (as in Articulatory Phonology) is chosen for the syllable in both a morphologically complex phonological word and a simplex word (e.g.

[bat]-er vs. *batter*). Thus, the phonetic encoding does not convey any information of the morphological structure within a phonological word. All these models would predict no effects of morpheme boundaries on phonetic implementation, but the data from the current study will show that this may not be true.

Overview

This paper reports two experiments, one using electromagnetic midsagittal articulography (EMA) and the other using electropalatography (EPG). Experiment I (using EMA) tests the difference between gestures separated by different morphological boundaries in Korean (lexicalized vs. nonlexicalized compound boundaries, and no boundary vs. a morpheme boundary). Experiment II (using EPG) examines effects of morpheme boundaries on palatalization in Korean, and explores the extent to which degree of palatalization varies as a function of the morphological affiliation of the gestures involved. While EMA gives us a reliable measurement of the temporal movements of articulators, it does not detect spatial movement outside the midsagittal plane. For this reason, EMA was used only for the investigation of the temporal relationship between gestures. EPG, on the other hand, cannot detect direct movements of articulators beyond the linguopalatal contact and thus is not the best tool for investigating the temporal relationship between gestures [see Byrd, 1994, 1995, 1996b, for the possibility of examining the temporal relationship using EPG], but it gives us a reasonable approximation of linguopalatal contact, from which the degree of palatalization, for example, can be examined [cf. Zsiga, 1993, 1995].

Both experiments will demonstrate that phonetic realization is affected by morphological structure. In a nutshell, I will show that intergestural timing is more stable inside a single morpheme than across a morpheme boundary, and also more stable across a lexicalized compound boundary than across a nonlexicalized compound boundary. In accounting for the stability of intergestural timing within a lexical item (a morpheme or a lexicalized compound), several hypotheses will be discussed including the Phase Window model [Byrd, 1996a], Bonding Strength [Browman and Goldstein, 1998], Phonological Timing [Zsiga, 1997] and Extended Phase Window model, newly proposed here. All of these models depart from the earlier model of Articulatory Phonology, in that they all provide some kind of mechanism to allow contextually induced gestural timing variability.

Before turning to the experiments, I will give a brief overview of these hypotheses. Full discussion of them will appear later in the ‘Discussion’ section. In Articulatory Phonology, the mechanisms for coordinating gestures are phasing rules, which state that a phase in one gesture is synchronized with a phase in another gesture. However, Byrd [1996a] points out that phasing rules do not capture the degree of overlap, which can vary depending on other linguistic and extralinguistic factors. In order to capture such contextual variability, Byrd [1996a] proposed the Phase Window model in which there are two ways of allowing variability in timing: one is through different phase (or timing) windows, i.e., a varying range of possible timing relationships, and the other is through weighting the phase window differently. In this model, any suprasegmental effects such as the effect of morpheme boundaries on timing should be dealt with by weighing the phase window differently, rather than by varying the width of the phase window. In the Extended Phase Window model, proposed in the current

Table 1. Two sets of speech materials for recording

a Set 1: Compound words

Lexicalized compounds	Nonlexicalized compounds
/ha**k-p**i/ 'tuition' ('study' + 'expenses')	/ha**k-p**i/ 'school corruption' ('school' + 'corruption')
/pɛ**k-p**al/ 'gray hair' ('white' + 'hair')	/pɛ**k-p**al/ 'white foot' ('white' + 'foot')
/t͡ʃʰə**p-t͡ʃ**a/ 'spy' ('spying' + 'person')	/t͡ʃʰə**p-t͡ʃ**a/ 'second wife's son' ('second wife' + 'person/son')

b Set 2: Mono- and heteromorphemic sequences

Monomorphemic /pi/	Heteromorphemic /p-i/
/na**pi**/ 'butterfly'	/na**p-i**/ 'lead' + Nom.
/sa**pi**/ 'private expense'	/sa**p-i**/ 'shovel' + Nom.
/t͡ʃa**pi**/ 'generosity'	/t͡ʃa**p-i**/ 'to handle' + 'nominalizer'

study, we will present the possibility that it is the phase window itself that is narrower for gestures within a lexical item (or a morpheme) than for those across lexical items. Results will also be discussed in terms of a few other alternatives. For example, Browman and Goldstein [1998] suggest that the variability in gestural timing can be handled by associating every phase relation within a lexical entry with a 'bonding strength' or the degree of cohesion between gestures involved: that is, the greater the bonding strength, the more stable the gestural timing. Another alternative is proposed by Zsiga [1997], who suggests that timing stability comes from categorical timing relations in phonology which are indicated by the presence or absence of association lines.

While these hypotheses may help us to understand the greater stability of timing found for a single lexical item as opposed to that found across different lexical items, it should be noted that none of them have been developed fully enough to make distinct predictions. Thus, in this paper, we will simply discuss how the results could be accounted for within each framework, in the hope that this will facilitate future research about gestural timing.

Experiment I

Method

In order to test the hypothesis that the variability of intergestural timing is affected by morphological boundaries, I compared sequences of segments in two different sets of conditions: (1) gestures separated by a lexicalized compound boundary vs. gestures separated by a nonlexicalized compound boundary (i.e. a word boundary) and (2) tautomorphemic gestures (i.e. no boundary) vs. gestures separated by a boundary of productively used morphemes. Two sets of speech materials are given in table 1.

In table 1a, a two-consonant sequence (e.g. *kp*) occurs both in a lexicalized compound word (e.g. *pekpal* 'gray hair') and in a nonlexicalized compound word (*pekpal* 'white foot'). The lexicalized

Table 2. Carrier sentences with test words

a Set 1: Compound words				
(*ha**kp**i* = 'tuition' or 'school corruption')				
nə	*ha**kp**i-etɛhɛ*	*anɨnkəs*	*is'ni?*	
you	____ about	to know-Rel thing	to have-Q	
'Did you hear about *"hakpi"*?'				
(*pɛ**kp**al* = 'gray hair' or 'white foot')				
kɨ	*salam*	*pɛ**kp**al-ine*		
that	person	_____ to be		
'That person has *"pɛkpal"*'				
(*t͡ʃʰə**pt͡ʃ**a* = 'spy' or 'son of the second wife')				
kɨ	*salam*	*t͡ʃʰə**pt͡ʃ**a-ja?*		
that	person	______ to be-Q		
'Is that person a *"t͡ʃʰə**pt͡ʃ**a"*'				
b Set 2: Tautomorphemic and heteromorphemic sequences				
nə _______	*p'alli hepolɛ?*			
you	quickly try to say-Q			
'Can you say _____ quickly?'				

compounds have meanings which are in general noncompositional and they are listed in dictionaries. On the other hand, nonlexical compounds may be viewed as noun phrases and their meanings are completely compositional. Comparing lexicalized and nonlexicalized compounds allows us to compare intergestural timing in one lexical entry and across two lexical entries.[1]

In table 1b, a *pi* sequence occurs both in a monomorphemic word (left column) and in an affixed form (right column). The affixed forms consist of a stem plus either a nominative suffix *i* or a nominalizer *i*. This set of words was used to examine the variability of intergestural timing in monomorphemic gestural sequences, compared with heteromorphemic ones (spanning a morphemic boundary). The matched segment sequences in different words ensure that any variability difference to be examined is not due to transition probabilities. It should be also noted that the items in tables 1a, b are segmentally different, so a comparison across separate sets (i.e. segments in a lexicalized compound, vs. segments in a tautomorphemic word) cannot be made.

The target words in table 1a were recorded in the carrier sentences in table 2a, and the target words in table 1b, in a frame sentence as in table 2b.[2] Each sentence was read with 15 or 16 repetitions from a randomized stack of cards, on which carrier sentences were written along with small iconic pictures to cue the intended target words. Speakers were told to read stimuli as comfortably as possible at their normal speech rate.

An EMA system (Carstens Articulograph AG 100) was used to track articulatory movements of tongue tip, tongue body, and lower lip. In the EMA system, three transmitter coils set up an electro-

[1] The meaning of lexicalized and nonlexicalized compounds can be disambiguated by the discourse context. So far, no apparent phonetic cues are known that might help disambiguate them. The current study tests the difference in gestural timing between them, and indeed shows that there is some phonetic difference. It would be interesting to test whether listeners can tell the difference without the help of context.

[2] In table 2, set 1 shows some inconsistency in that the carrier phrases use different intonational patterns (questions and declarative sentences) and that the location of the target item in the sentence varies. This should not be problematic, however, because the comparisons to be made occur only between two homophonous items occurring in an identical context.

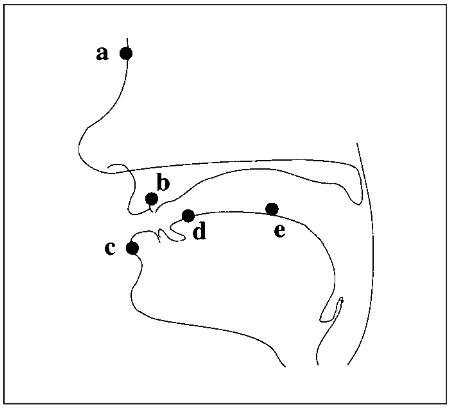

Fig. 2. Five locations of transducer coils, including two reference points (a = nose, b = upper front gum line) and the three articulators (c = lower lip at the vermilion border, d = tongue blade, e = tongue body).

magnetic field around the head of a speaker. Within that field, the horizontal and vertical movements of receiver (transducer) coils attached midsagitally to articulators (e.g. lips or tongue) can be tracked. In this experiment, five transducer coils were attached: one as two reference points to (a) the nose and (b) upper gumline; three to the articulators – (c) lower lip at the vermilion border, (d) tongue blade about 1 cm from the tongue tip, and (e) tongue body about 4–4.5 cm from the tongue tip, as shown in figure 2. The coils at the lower lip, tongue blade and tongue body track kinematics of the labial gesture for /p/, tongue tip/blade gesture for /t͡ʃ/ and the tongue body gesture for /k/, respectively. (For more technical information on the Carstens System see Schoenle [1988], Schoenle et al. [1989], Tuller et al. [1990]; see also Perkell et al. [1992] for a different articulograph system.)

Three speakers of Seoul Korean (H.L., H.H., and T.C., the author) participated in recording for the first experiment. Both acoustic and EMA recording were made with sampling rates of 16 kHz and 665 Hz (1.5-ms interval), respectively. However, when the acoustic data were synchronized with the EMA data, they were compressed by the speech processor into a 4 bit ADPCM format with reduced frequency range of a maximum of 7 kHz. Therefore, only basic acoustic segmentations were made (e.g. between a vowel and a voiceless stop) from the acoustic waveform. (With this sampling rate, 7 kHz, there is about 0.15-ms interval between samples, which is small enough to detect the boundary between a vowel and a completely voiceless stop.)

Measurement

The data were analyzed by the Carstens program, Emalyse 3.0, which makes it possible to measure some key properties of articulatory events, such as velocity maximum points of the closing and releasing gestures of consonants and the peak vertical displacement for each consonant in the cluster.

One way to investigate the variability in timing for a consonant sequence is to examine standard deviations for measured intervals between two articulatory defined events, such as C-centers, as illustrated in figure 3 [Browman and Goldstein, 1988; Byrd, 1995]. In the figure, the C-center is the midpoint of the plateau for each consonant.[3] Browman and Goldstein [1988], using the vertical displacement trace of an X-ray microbeam pellet attached to an articulator, defined the plateau as the region of the curve within about 1.3 mm from the peak. Analogously, the plateau in the present EMA study was defined as the region of the curve within about 1.3 mm of displacement from the peak. The left and

[3] In Browman and Goldstein [1988], the C-center was the midpoint of plateaus of one or more consecutive consonantal gestures. It was originally used for the timing relation between the consonantal and vocalic gestures. But in this paper, I adopt C-centers to examine the timing relationship between consonantal gestures.

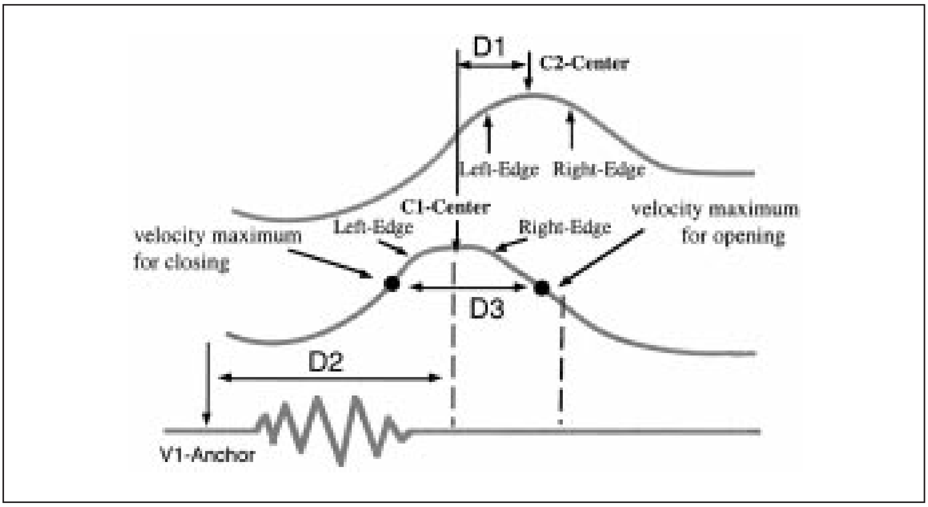

Fig. 3. Definition of measurements of articulatory movement of C_1 and C_2 in a cluster.

right edges are the onset and offset points of the plateau. Browman and Goldstein [1988] define the V_1-anchor point as the acoustic midpoint of the consonant which precedes the vowel, but in the present study, the V_1-anchor point was defined as the articulatory midpoint of the plateau of the preceding consonant, rather than the acoustic midpoint, because lenis stops become voiced intervocalically, making segmentation inconsistent. Note also that the target word *hakpi* was excluded for this measure because no midpoint for /h/ was detectable due to intervocalic voicing.

Lexicalized vs. Nonlexicalized Compounds

In the lexical vs. nonlexicalized compounds, the test segment sequence consisted of two stops. First, intervals were measured between C-centers of adjacent consonants in the clusters (i.e. *pk* in *pekpal* and *hakpi* and *pt͡ʃ* in *t͡ʃʰəpt͡ʃa*, D1 in fig. 3). (It was not possible to make this measurement for *hakpi* produced by speaker H.H., who made no separable dorsum movements for *k* and the following *i*.) Then, standard deviations for D1 were taken as an index of variability of the temporal coordination for the two consonants in a sequence.

In addition, two other measurements were taken in order to examine the extent to which any observed variations in intergestural timing are accounted for by variations in duration of segments preceding a morphological boundary. The first was the interval between $V_{1\text{-anchor and }C}$1-center in $V_1C_1\#C_2V_2$ clusters (D2 in fig. 3). The second was the duration of the closing (velocity maximum to peak) and opening (peak to velocity maximum) of the C_1 gesture (D3 in fig. 3).

It is possible that in the environment of $V_1C_1\#C_2V_2$ (where # refers to a morpheme boundary inside compounds), V_1C_1 is subject to final lengthening because of the morphological boundary. If there is final lengthening of V_1 in $V_1C_1\#C_2V_2$ in a nonlexical compound, intervals between V_1-anchor and C_1-centers (D2, which is an indirect measure of duration of the vowel gesture) will be increased; if there is final lengthening of C_1, durations of closing and opening movements for the C_1 (D3) gesture will be increased. Changes in C_1 duration affect not only D2 and D3, but also D1 (the interval between C_1-and C_2-centers) since D1 too involves part of the C_1 duration. Thus, D1 is expected to be correlated to some extent with D2 and D3. For these durational measurements, the target word ***hak-pi*** was excluded, because the duration of the opening gesture of /k/ could not be measured due to its coarticulation with the following high vowel /i/.

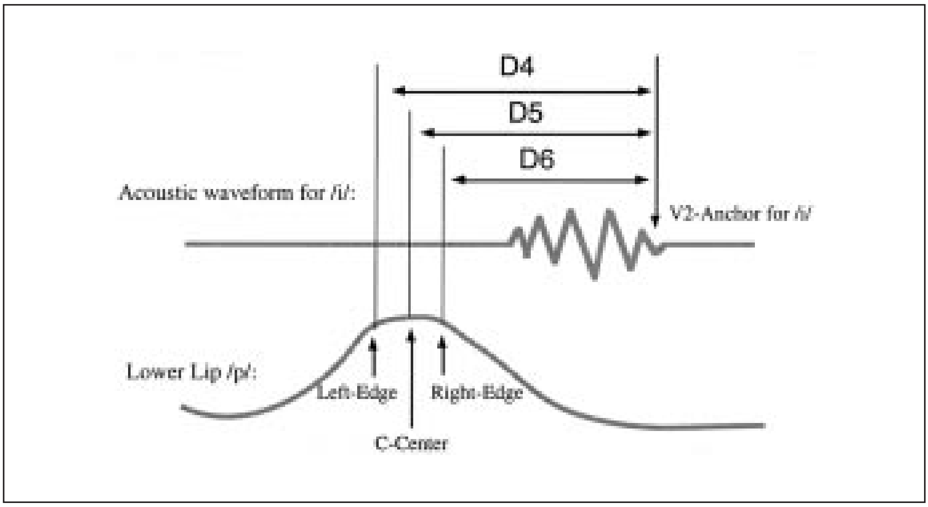

Fig. 4. A schema of the sequence *pi* showing the left edge (the onset point of the plateau), C-center (the temporal center of the plateau), right edge (the offset point of the plateau) for a consonant [p] and the V_2-anchor of a following vowel [i].

Monomorphemic pi *vs. Heteromorphemic* pi

In the monomorphemic vs. heteromorphemic condition, the test segment sequence was CV. The variability in timing organization between a consonant and a vowel (i.e. *pi*) can be also assessed from standard deviations of measured intervals between articulatorily defined points for the consonants (i.e. C-center, right edge, left edge) and acoustically defined point for the vowel (i.e. V_2-anchor, the end-point of the acoustic periodic waveform[4]) as illustrated in figure 4. The same method was employed in Byrd [1995], and Browman and Goldstein [1988]. D4 is the interval between the left edge and the V_2-anchor; D5 is the interval between the C-center and the V_2-anchor, and D6 is the interval between the right edge and the V_2-anchor. Then, as an index of variability, standard deviations for D5 were taken. D5, rather than just the acoustic vowel duration, was chosen to measure vowel variability because the onset of the vowel gesture usually occurs around the midpoint of the preceding bilabial consonant [Byrd, 1995; Browman and Goldstein, 1988].

The Levene F Test

While standard deviations were used as an index of relative variability between two conditions, the Levene F test [Levene, 1960; Dixon, 1988, cited in Byrd, 1994, 1996b] was employed to test for equality of variance. The Levene statistic allows us to compare variability within pooled data [see Byrd, 1996b] since it compares the deviations of each data point from the group mean. For each speaker, the data were separated for each group (e.g. lexicalized and nonlexicalized compounds) and the mean was computed independently for each group. Then, the absolute difference between the data points and their respective group means was calculated. Next, the pooled deviation data were submitted to a repeated measures ANOVA (the Levene F statistic), with the Speaker variable as random inde-

[4] Although the acoustic resolution was not very high, the endpoint of the acoustic periodic waveform for the vowel *i* was located easily because of the complete voicelessness of the following fortis stop in the carrier sentence. Note that the fortis stops never become voiced, so that the discontinuity of voicing is located easily. Note also that V_2-anchor could have been defined articulatorily. But in order to be comparable with the second experiment, an EPG study in which defining V/$\widehat{t\int}$-anchor articulatorily was impossible, an acoustically defined V_2-anchor was chosen. The validity of this point as the vowel anchor is discussed in Byrd [1995].

Table 3. Comparisons of standard deviations of C_1C_2 sequences between lexicalized and nonlexicalized compounds (measures in milliseconds)

	Speaker H.L.			Speaker T.C.			Speaker H.H.		
	lexical CC		nonlexical CC	lexical CC		nonlexical CC	lexical CC		nonlexical CC
pɛ**k-p**al	10.25	<	12.94	4.29	<	9.11	8.98	<	12.59
	(37.95)		(43.12)	(31.75)		(36.10)	(14.70)		(15.16)
t͡ʃʰə**p-t**͡ʃa	6.73	<	8.56	5.21	<	12.33	5.10	<	9.28
	(48.6)		(51.2)	(46.5)		(60.2)	(26.5)		(27.3)
ha**k-p**i	6.36	<	9.30	6.27	<	12.84	–		–
	(39.7)		(55.1)	(30.4)		(32.6)			

Means are provided in parentheses. For H.H., data for *hakpi* is not available due to dorsum movements not separable between *k* and *i* (n = 15 or 16 for each test item).

pendent variable. The Levene F test using the repeated measures ANOVA allows us to determine if there is a significant difference in variability between the conditions (the lexical vs. nonlexical compounds and the monomorphemic vs. heteromorphemic sequences) and between the speakers. StatView 5.0 (SAS Institute Inc., 1998) was used to perform the statistical tests.

Results

Nonlexicalized vs. Lexicalized Compounds

Let us first examine the standard deviations obtained from the interval between C_1 and C_2 gestures (D1 in fig. 3) in set 1. As summarized in table 3, for all 3 speakers, the standard deviations were always greater for $C_1\#C_2$ in nonlexicalized compounds than in lexicalized compounds.

This is true for all three different pairs of compound words tested. Variations in timing in $C_1\#C_2$ are shown in figure 5. In the figure, intervals between C_1 and C_2, from which standard deviations of D1 were obtained, were plotted according to percentiles of the variable interval. For each box plotted on the X axis there are 5 horizontal lines, which represent the 10th, 25th, 50th (median), 75th and 90th percentiles of the variable, respectively. The empty dots are values either above the 90th or below the 10th percentiles of the variable. In the figure, values for nonlexicalized compounds (white boxes) are more widely distributed in the range not only between the 10th and the 90th percentiles but also in the range between the 25th and the 75th percentiles, especially for speakers T.C. and H.H., as compared to lexicalized compounds (gray boxes). The wider range of data, along with greater standard deviations associated with nonlexicalized compounds, suggests that intergestural timing is more variable for the C_1C_2 sequence of nonlexicalized compounds compared to that of lexicalized ones.[5]

[5] Figure 5 also shows differences in median values between lexicalized and nonlexicalized compounds. Lexicalized compounds tend to have shorter intervals between C-centers than do nonlexicalized compounds. One may suspect that the observed tendency towards shorter intervals may be the result of the greater gestural overlap. However, we cannot measure the degree of gestural overlap directly from durations between two C-centers: in Articulatory Phonology, the degree of overlap is a relative percentage, while figure 5 is in milliseconds.

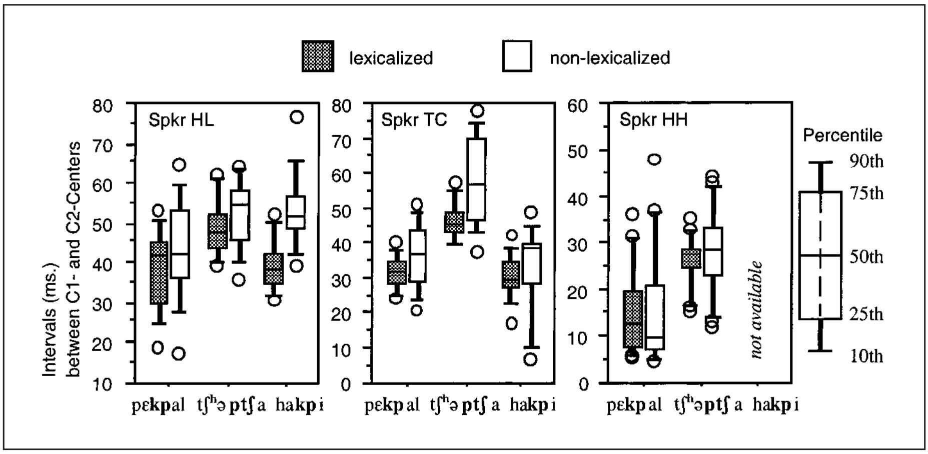

Fig. 5. Box plot displaying the distribution of intervals (ms) between C-centers. Error bars refer not to the standard deviation but to the distribution of the values between the 10th and the 90th percentiles. The empty circles in the plot are values above the 90th and below the 10th percentile.

Most of these visual observations of the difference in variability were confirmed by a Levene F test. A repeated measures ANOVA, using the absolute values of deviations from the mean, showed that there was a significant difference between lexicalized and nonlexicalized compounds for two test words, *t͡ʃʰə**pt͡ʃ**a* and *ha**kp**i* [for *t͡ʃʰə**pt͡ʃ**a*, $F_{(1,\ 2)} = 12.49$, $p = 0.0016$; for *ha**kp**i*, $F_{(1, 2)} = 4.52$, $p = 0.0440$] with no significant speaker effect. For *kp* in *pe**kp**al*, there was only a trend towards greater variability in nonlexicalized compound [$F_{(1, 2)} = 4.15$, $p = 0.0511$]. However, when we excluded speaker H.L., who showed the least difference in standard deviation, the difference was significant [$F_{(1, 1)} = 5.19$, $p = 0.03$].

Monomorphemic vs. Heteromorphemic Sequences of [pi]

Table 4 shows standard deviations of D5 (fig. 4), the interval between C-center and V-anchor of *pi* sequences.[6] For all 3 speakers, the standard deviations are greater for heteromorphemic sequences than for monomorphemic sequences, except for H.L.'s *t͡ʃapi*, which shows no obvious difference in standard deviations.[7] The range of intergestural timing variations is shown in figure 6. The box plot in figure 6 shows a wider range of values for heteromorphemic sequences than tautomorphemic sequences

[6] The differences due to the presence or absence of a morpheme boundary in table 4 are greater than differences due to the presence or absence of a compound boundary in table 3. Such differences are likely due to the fact that timing relations of CC sequences are generally more stable than CV gestural timing relations. This is consistent with findings that, as compared to consonants, vowels tend to be more sensitive to change in speech rate [e.g. Gay et al., 1974; Miller, 1981; Flege, 1988]. This also may explain why the standard deviations in CC sequences in table 3 are generally smaller than those in CV sequences in table 6 in experiment II, despite the fact that the CV sequences belong to the same word, while the CC sequences belong to different words.

[7] The affix-*i* in the heteromorphemic *t͡ʃapi* is a nominalizer while the affix-*i* in other heteromorphemic words is a nominative marker. We gather that the lack of difference in standard deviations for H.L. and the relatively smaller difference for H.H. for *t͡ʃapi* compared to *sapi* or *napi* is consistent with claims in the literature that derivational morphology is more likely listed in the lexicon than inflectional morphology.

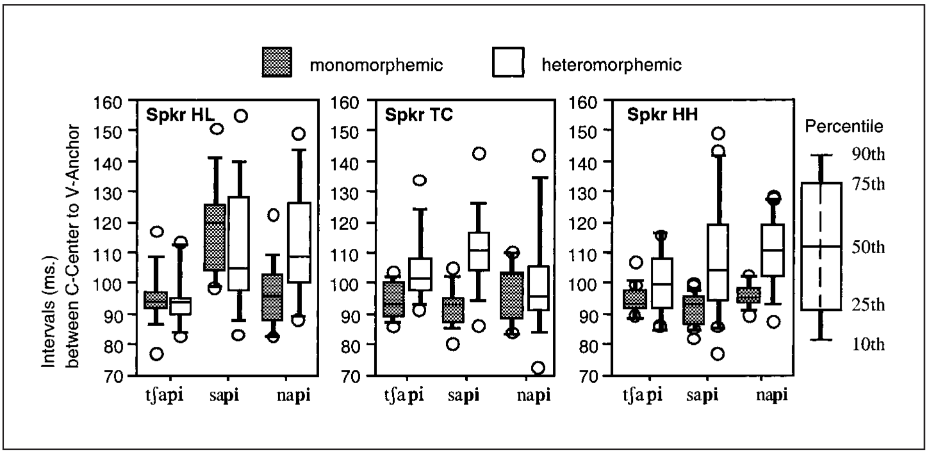

Fig. 6. Box plot displaying the distributions of intervals (ms) from C-center to V-anchor in *pi*. Error bars refer not to the standard deviation, but to the distribution of the values between the 10th and the 90th percentiles. The empty circles in the plot are values above the 90th and below the 10th percentile.

Table 4. Comparisons of standard deviations of *pi* sequences between tautomorphemic and heteromorphemic words (measures in milliseconds)

	Speaker H.L.			Speaker T.C.			Speaker H.H.		
	tauto-morphemic		hetero-morphemic	tauto-morphemic		hetero-morphemic	tauto-morphemic		hetero-morphemic
t͡ʃa**pi**	8.99	<	8.96	8.51	<	12.68	4.53	<	10.31
	(94.8)		(94.5)	(94.2)		(104.4)	(93.8)		(100.7)
sa**pi**	15.61	<	20.03	6.40	<	14.66	4.99	<	19.97
	(118.1)		(110.7)	(92.7)		(111.2)	(81.9)		(112.6)
na**pi**	11.52	<	19.08	3.52	<	17.72	3.25	<	12.14
	(95.6)		(106.1)	(98.4)		(101.2)	(92.2)		(114.4)

Means are provided in parentheses (n = 15 or 16 for each test item).

including *t͡ʃapi* for H.L. A Levene F test showed that the difference in variability is significantly reliable for all test words [for *t͡ʃa**pi***, $F_{(1, 2)} = 10.47$, $p = 0.0033$; for *sa**pi***, $F_{(1, 2)} = 17.26$, $p = 0.0003$; for *na**pi***, $F_{(1, 2)} = 15.49$, $p = 0.0006$]. There was no significant difference between speakers, despite the fact that the difference in standard deviations for *t͡ʃapi* for H.L. was small.

Thus far we have relied on standard deviations and Levene statistics to examine effects of morpheme boundaries on timing of gestures. One caveat should be addressed: as shown in tables 3 and 4, the group means are generally smaller for the monomorphemic sequences and lexicalized compounds than for the heteromorphemic

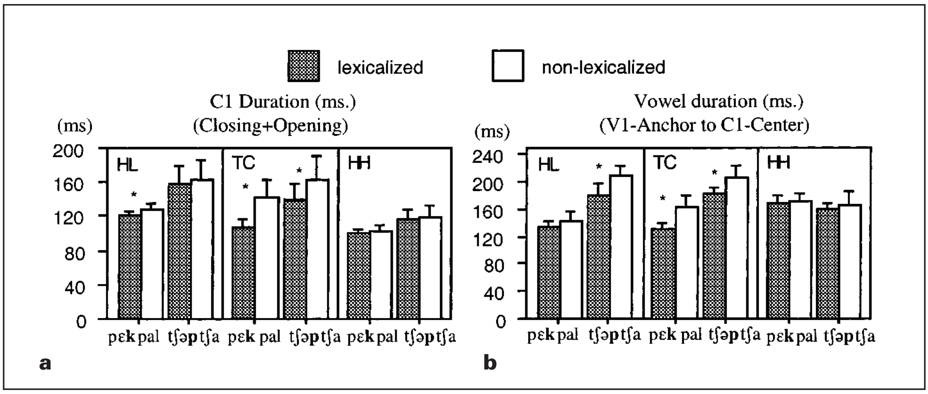

Fig. 7. Duration of gestures for C_1 in $V_1C_1\#C_2V_2$. **a** Articulatory duration of C_1 (with closing and opening gestures combined). **b** Approximated articulatory duration of V_1. The error bars indicate standard deviations. (An asterisk refers to significant difference between lexicalized and nonlexicalized compounds at $p < 0.05$ obtained from t test.)

sequences and nonlexicalized compounds. Variances may be scaled proportionally with means in some cases especially when data are distributed unevenly (or 'heteroscedastically') across the range of the variable. While it may be possible that the greater standard deviations found in this study are in part due to the heteroscedasticity, this cannot be the only source of variation. One piece of evidence is that, in four out of eight comparisons and three out of nine comparisons in tables 3 and 4, respectively, the difference in standard deviations between groups that are compared is far greater than the difference in means. Even if variances were scaled directly with means, difference in standard deviations could not exceed differences in means, given that standard deviation is a square root of variance. Another piece of evidence can be found in the Levene statistics using not the absolute but normalized deviation from the group mean, scaled proportionally to the mean, i.e. percent deviation relative to the mean. This normalized deviation should minimize the potential dependency of variance on means. In fact, results of the Levene statistics using the normalized values show significant effects in most cases in which the Levene statistics using the absolute deviation values show significant effects. For example, for the tautomorphemic vs. heteromorphemic condition, significance was found for all pairs of test words [for *t͡ʃ**api***, $F_{(1, 2)} = 8.365$, $p = 0.0076$; for *s**api***, $F_{(1, 2)} = 10.1$, $p = 0.0038$; for *n**api***, $F_{(1, 2)} = 16.1$, $p = 0.0004$]. However, for lexical vs. nonlexicalized compounds, significance was found for *t͡ʃʰə**pt͡ʃ**a* [$F_{(1, 2)} = 6.45$, $p = 0.0146$], but not for *ha**kp**i* [$F_{(1, 1)} = 3.35$, $p = 0.0786$] for which significance was found at the level of $p < 0.05$ in the Levene test using the absolute deviations. Considering all these together, we can conclude that the variability difference found in this study is not primarily due to the mean and variance dependency, although variance may be correlated in part with means.

Final Lengthening and Nonlexicalized Compounds

In general, there is considerable speaker variation, with an overall tendency towards longer C_1 in nonlexicalized compounds compared to lexicalized ones. This is shown in figure 7a. As can be seen, speaker T.C. shows a significant main effect of the lexical status on articulatory duration of C_1 for each test word [$t_{(28)} = -5.300$, $p < 0.0001$ for *pɛ**k-p**al;* $t_{(28)} = -2.704$, $p = 0.0119$ for *t͡ʃʰə**p-t͡ʃ**a*]; speaker H.L. shows a significant difference just for *pɛ**k-p**al* [$t_{(28)} = -2.981$, $p = 0.0059$]; speaker H.H. shows no significant difference at all.

Similar patterns in vowel gesture lengthening were found as shown in figure 7b. For T.C., the duration of the vowel gesture, indicated by the interval between the V_1-anchor and the C_1-center in $V_1C_1\#C_2V_2$, is significantly longer in nonlexicalized compounds than in lexicalized ones [$t_{(28)} = -6.187$, $p < 0.0001$ for *pɛ**k-p**al;* $t_{(28)} = -3.954$, $p = 0.0005$]; for H.L., the vowel is significantly lengthened for *t͡ʃʰə**p-t͡ʃ**a* [$t_{(28)} = -4.736$, $p < 0.0001$] but not for *pɛ**k-p**al.* As in C_1 duration, for H.H., no significant difference was found for either of the test words.

Now, let us turn back to our original question of whether the observed final lengthening explains the intergestural timing variation between consonants which span a morpheme boundary. Combining the results above, the difference in durational lengthening of V_1C_1 does not seem to be a consistent factor that differentiates lexicalized and nonlexicalized compound words, but rather a speaker-specific phenomenon. In other words, while speakers in general showed significant differences in timing variability, only 1 speaker (T.C.) shows a significant durational difference for both test words, while the other 2 speakers show significant differences for only one test word (H.L.) or none (H.H.). This asymmetric result suggests that the greater variations in intergestural timing for nonlexicalized compounds do not come entirely from the observed final lengthening of either V_1 or C_1 preceding a morphological boundary. Nonetheless, it should be noted that our data clearly suggest that the variability in timing is correlated with the word-final lengthening to some degree. Of course, it is not clear how much of the observed variability is accounted for by the lengthening effect as opposed to the difference in CC timing relationship. It is equally unclear whether these two effects are after all separable: it may be the case that the lengthening of C_1 occurs precisely because the CC timing relation is less stable. What emerges from the data in any event is that intergestural timing for nonlexicalized compounds is more variable than for lexicalized compounds.

Summary of Experiment I

To sum up, the results of the EMA study show that any gestural sequence within a single lexical entry (including lexicalized compounds and monomorphemic words) shows more stable intergestural timing patterns on the surface as compared to a gestural sequence across different lexical entries. Specifically, intergestural timing is more stable for two sequential consonant gestures spanning a morphological boundary inside a lexicalized compound than inside a nonlexicalized compound; intergestural timing between consonant and vowel gestures inside a word was also found to be more stable when tautomorphemic than when heteromorphemic. In the EPG study that follows, we will further investigate spatial and temporal characteristics of CV sequences in Korean palatalization.

Experiment II

EPG provides a useful tool for comparing different patterns of palatalization [Barry, 1992; Zsiga, 1993]. Although some of the methodological details employed in the present study are different from those found in Barry [1992] and Zsiga [1993], the fundamental idea of their EPG studies provides the tools to investigate the spatial and temporal relations between adjacent gestures. In this experiment, palatalization is interpreted as the temporal overlap of gestures for a coronal consonant (e.g. *t, n*) and a following vowel *i*, adopting the notion in Articulatory Phonology that assimilations are due to increased overlap between gestures.

In Korean, palatalization occurs when a coronal consonant *(t, n)* is followed by a vowel *i* [e.g. Kim-Renaud, 1974; Kim, 1976; Ahn, 1985; Iverson and Wheeler, 1988]. However, *t* palatalization occurs only when the triggering vowel is not in the same morpheme as the target consonant *t*, while *n* palatalization occurs before *i* regardless of the environments. This asymmetry between *t* and *n* palatalizations is illustrated below. (Note that obstruents become [+ voice] intervocalically by an independent process.)

(a)	/k'ini/	→	[k'iɲi]	'meal'
(b)	/san-i/	→	[saɲi]	'mountain-Nom'
(c)	/mati/	→	[madi]	'knot'
(d)	/mat-i/	→	[mad͡ʒi]	'the eldest'

While *n* palatalization occurs regardless of the environment (a, b), *t* is not palatalized in the nonderived, tautomorphemic, environment (c). The failure of *t* palatalization morpheme-internally can be seen as an instance of 'nonderived environment blocking' [e.g. Kiparsky, 1973].[8]

The findings in experiment I make a prediction about the historical development of palatalization in derived vs. nonderived environments: we should expect that, before palatalization of coronal stops in front of /i/ became a phonological rule, intergestural timing between a coronal and a following vowel *i* was considerably less variable for *nonderived* monomorphemic sequences (in a single lexical entry) than for *derived* heteromorphemic sequences (across different lexical entries). If this was the case, then the gestures with less stable timing as in /t-i/ would have been subject to historical changes due to varying gestural overlap, e.g. palatalization. This prediction is in spirit of a hypothesis made in Articulatory Phonology that gestural 'sliding' which results in 'perceptual' deletions of consonants or assimilations in casual speech [e.g. Browman and Goldstein, 1990] is most likely to occur at the end of one lexical unit and the beginning of the next one, a position in which gestures are less cohesive [Browman and Goldstein, 1998].

In addition, this experiment examines the extent to which degree of palatalization differs for the target consonants *t* and *n*. According to a traditional assumption, the heteromorphemic or derived *t-i* sequence is expected to undergo palatalization, but the monomorphemic *ti* is not. However, there are two competing predictions about *n-palatalization. Under the traditional assumption [Ahn, 1985] that* n-palatalization is postlexical, we would predict that intergestural timing and degree of gestural overlap for *n* will not vary depending on the morphological or lexical environment. On the other hand, if we assume that the lexical status of intergestural timing affects *n*-palatalization, then intergestural timing should be less variable for tautomorphemic *ni* than for heteromorphemic *n-i,* just as for *ti* vs. *t-i.*

[8] However, more recently Kiparsky [1993] has proposed that all coronal consonants (including *t*) undergo palatalization regardless of the environment, derived or not. Based on impressionistic observations, Kiparsky argues that the only difference between /ti/ and /t-i/ is whether frication is involved or not. So, in his transcription, /ti/ is realized as [tʲi], and /t-i/ as [tʃʲi]. The current study, however, shows that only /t-i/ is palatalized, while /ti/ is not, contra Kiparsky's observation.

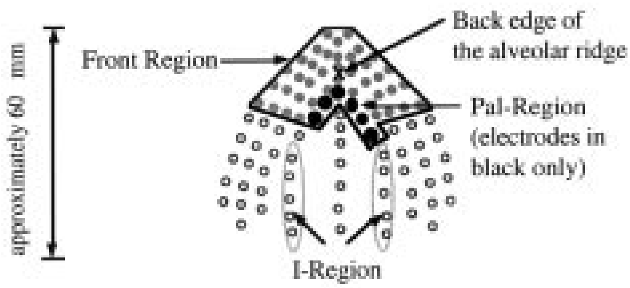

Fig. 8. Three defined regions. The front region is roughly equivalent to the dental, alveolar, and postalveolar areas combined; the Pal-region is the area that is almost always contacted for the production of underlying palatal consonants; the i-region is a subset of the contact area for the vowel *i* in *papi* (see fig. 9 for detail).

Table 5. Test words for experiment II

Test words			
pa**ni**	'name'	ma**ti**	'knot'
pa**n-i**	'class-Nom'	ma**t-i**	'the oldest'
Carrier sentences			
nə pa**ni** pwakk'una	'you saw *pani*, right?'		
nə pa**n-i** makk'una	'you are in the right classroom, right?'		
nə ma**ti** pwakk'una	'you saw (that) knot, right?'		
nə ma**t-i** pwakk'una	'you saw the eldest, right?'		

Method

EPG data were recorded from 3 native speakers of Seoul Korean, 2 male speakers (N.L. and T.C., the author) and 1 female speaker (J.Y.). Speakers wore custom-made pseudopalates with 96 electrodes covering the entire hard palate and the inside surfaces of the upper teeth (Kay Elemetrics Palatometer 6300). When an electrode is contacted by the tongue, a circuit is completed and the contact is recorded by the Palatometer. The target words were repeated 14 times in a carrier sentence where segmental and prosodic environments were controlled for table 5. All four sentences were written in a single sheet of paper, repeated in 14 different sheets, each of which had a different order of sentences. Speakers were told to speak sentences as naturally as possible at a normal speech rate.

In order to examine the degree of gestural overlap between a consonant and a following vowel *i*, three separate contact regions were defined as schematized in figure 8. Six hypothetical words *(pat͡ʃi, papi, pata, pate, pana, pane)* in isolation were also recorded (10 repetitions each); these were used as a reference in determining the relevant contact regions. The first region is the linguopalatal contact area, which is assumed to be contacted by the tongue tip and/or the tongue blade, and covers the dental, alveolar and postalveolar regions. I will call this region the *front-region*, which includes any electrode that was contacted more than once for 10 repetitions of each consonant (/t/, /n/, /t͡ʃ/). The second region is the postalveolar palatal region, a subset of the front-region, in which no electrodes are contacted for anterior coronals (*n* or *t*) when followed by nonhigh vowels *a* and *e* but are contacted exclu-

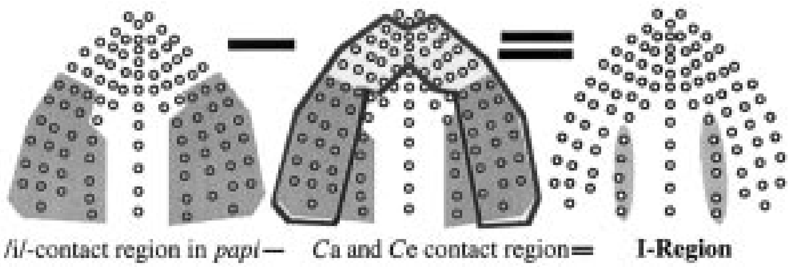

Fig. 9. Defining the i-region. The i-region is a subset of the contact region for the vowel *i* in *papi* that is never contacted during the production of *na, ne, ta, te, pa,* or *pe.*

sively for the underlying $\widehat{t\int}$ in *pa*$\widehat{t\int}$*i:* that is, when *t* or *n* in reference words *pata, pate, pana,* and *pane* are produced, the Pal-region is never contacted. In defining the Pal-region, any electrode contacted in this region more than once in 10 repetitions was included, but it turned out that each electrode was contacted at least 7 times or more out of 10 repetitions.

Note that the Pal-region covers the area starting from at least 3 mm behind the corner of the alveolar ridge (the front of which is in general considered as a landmark to distinguish [+ant] from [–ant] or alveolar from postalveolar [see Keating, 1991] to the electrodes that are contacted by the underlying postalveolar palatal /$\widehat{t\int}$/ in $\widehat{t\int}$*i* sequence. In order to include the rearmost possible electrodes, the vowel /i/, rather than a nonhigh vowel, was used as the following vowel. However, the defined Pal-region was never contacted during /i/ in a control case, *papi* (see below).

Finally, the third region is the *i-region,* defined as the area that was contacted by the tongue body when only the front high vowel *i* in the control word *papi* is produced. This area is *i*-specific and is never contacted when the mouth is sealed for the closure of *t* or *n* followed by a nonhigh vowel (i.e. *a* or *e*). The procedure for defining the i-region, which is similar for the Pal-region, is illustrated in figure 9. These regions, defined separately for each speaker, then were used to derive contact profiles from which articulatory movements were estimated.

Measurement

Following the method employed by Byrd [1995], EPG contact profiles for the front-region and the i-region were used as the basis for measures of articulatory kinematics for the tongue tip/blade and tongue body, respectively. As in the EMA experiment, standard deviations obtained from intervals between articulatory defined points of the consonant and the V-anchor point were used to investigate timing variability. The measurement points are illustrated in figure 10. Note that in the front-region, the plateau is defined as a change of two electrodes from the maximum contact point.

Standard deviations of three different intervals were measured, as described in M1.

M1: Interval standard deviations: standard deviations obtained from three different measured intervals: (1) from the left edge of a consonant (i.e. *t* or *n*) to the V-anchor of the following vowel *i,* (2) from the C-center to the V-anchor, and (3) from the right edge to the V-anchor.

Three different measured intervals were examined because we do not know how a Korean consonant-vowel sequence is temporally organized, or which point of the consonant has the most stable temporal coordination with the following V-anchor point. (In English, the C-center generally has the most stable relationship with the following V-anchor point [Browman and Goldstein, 1988;

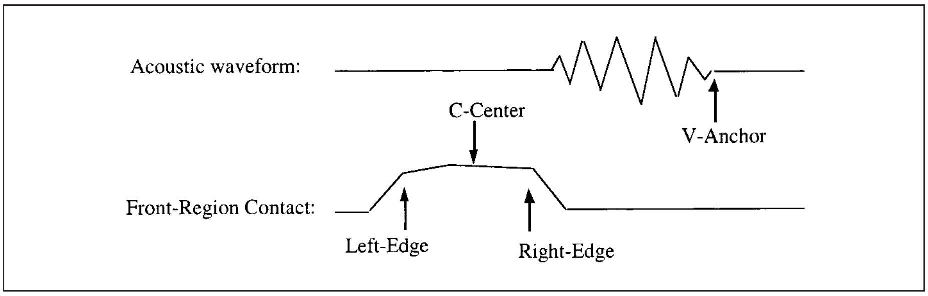

Fig. 10. Schema for a sequence of a coronal consonant and a vowel *i*, showing the left edge (the onset point of the plateau), C-center (the temporal center of the plateau), and right edge (the offset point of the plateau) for the consonant, and the V-anchor for the vowel.

Byrd, 1995].) In addition, three other measurements were made to examine the degree of intergestural overlap:

M2: Maximum contact in the Pal-region: the percent of maximum contact in the Pal-region.

M3: Maximum contact in the i-region: the percent of maximum contact in the i-region at the time when the contact in the front-region reaches the left edge.

M4: Sequence overlap (%): the percent of overlapping duration between the front-region and i-region compared to the total sequence duration.

The percent of electrodes contacted in the Pal-region and the i-region, respectively (i.e. M2 and M3), was measured at the point of maximal constriction, when the contact in the front-region reached the left-edge of the plateau [see Byrd et al., 1995, for methodology]. Palatalization can refer to any combination of three independent articulatory components: tongue fronting (or retracting), tongue raising, and spirantization [see Lahiri and Evers, 1991; Bhat, 1978]. The degree of tongue retraction and tongue raising can be observed by looking at the amount of contact in the Pal-region and the i-region, respectively. In addition, the percentage of the maximum i-region contact measured at the left edge of the plateau shows how much the articulatory gesture of the tongue dorsum for *i* has been anticipated (and is thus overlapped with the preceding consonant) at that point. To be more precise, the Pal-region contact can be the index of the gestural *blend* between a consonant (*t* and *n*) and the vowel *i*, both involving the coronal articulation, whereas the i-region contact at the left edge can be the index of the gestural overlap between tongue tip and tongue body gestures.

Finally, sequence overlap (%), M4, measures the degree of intergestural overlap between front- and i-regions, which is assumed to be another index of the degree of palatalization. A sample contact profile that is used for the measurement of sequence overlap (%) for the sequence of *ni* in *pani* is given in figure 11. (Note that the maximum contact for the i-region in figure 11 is about 60%. Recall that the control case used for defining the i-region was produced in isolation in order to elicit an unreduced form. By contrast, the target words tested were produced in a carrier sentence with a normal speech rate comfortable to the speaker. This accounts for why there is generally less contact in the i-region for the target words than for the control case.) In the figure, point A is the onset of the front-region contact; point B is the onset of the sequence overlap; point C is the offset of the sequence overlap, and point D is the offset of i-region contact. Sequence overlap (%) is the percent duration between point B and point C as compared with the entire sequence duration between points A and D.

The coded data were analyzed using a two-way analysis of variance. The factors considered were morphemic status (tautomorphemic vs. heteromorphemic) and consonant identity (*n* vs. *t*) for each of three different measurements (i.e. M2, M3, and M4) described above. As in experiment I, a repeated measures ANOVA with Speaker as a random variable was employed to examine the overall effect across speakers. When there was a significant speaker effect, post hoc comparisons were made separately for each speaker.

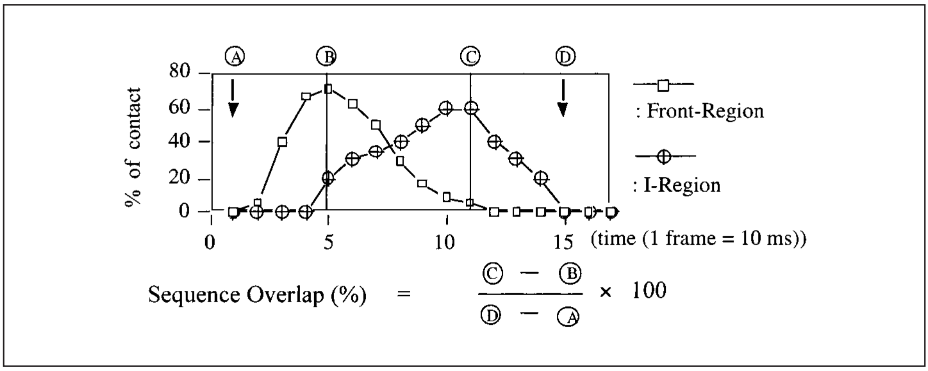

Fig. 11. Sample contact profile for *ni* sequence in *pani* and sequence overlap (%).

Results

Variability in Intergestural Timing

Let us first examine the standard deviations. As shown in table 6,[9] the standard deviations are almost always greater when the consonant and vowel gestures are heteromorphemic than when tautomorphemic. This is true for both *t* and *n*. The only exception is for the speaker J.Y.; for this speaker, the standard deviation for the tautomorphemic *ni* sequence is greater for the interval between the left edge to the V-anchor, but the interval between the right edge and the V-anchor has a greater standard deviation for the heteromorphemic *n-i* sequence, showing greater variability.

Overall, the smaller standard deviations for the tautomorphemic sequence in general imply that the intergestural timing between a coronal consonant (*t* or *n*) and a vowel *i* within the same morpheme is more stable. This is in agreement with the finding in experiment I that the intergestural timing within a single morpheme is more stable than across a morpheme boundary. Another fact that should be emphasized here is that a greater variability was found in not only the heteromorphemic *t-i* sequence but also in the *n-i* sequence, when compared to their tautomorphemic counterparts.

In order to test the significance of these numerical differences, a Levene test was once again employed, using a two-way ANOVA with the morphemic status and consonant identity (*n* vs. *t*) as independent variables. There was an overall significant difference in variability between tautomorphemic and heteromorphemic sequences for the intervals between the C-center and the V-anchor [$F_{(1, 2)} = 4.116$, $p = 0.0477$] and between the right edge and the V-anchor [$F_{(1, 2)} = 5.451$, $p = 0.0235$]; for the interval between the left edge and the V-anchor, there was also a trend towards greater variability for the heteromorphemic sequence [$F_{(1, 2)} = 3.644$, $p = 0.0619$]. On the other hand, there was no effect of consonant identity, and no interactions between the factors for

[9] The results shown in table 6 indeed suggest that the C-center may be most stably coordinated with the following V-anchor point in Korean, as in English. Overall, standard deviations are smaller for the C-center measures than for the left edge or the right edge.

Table 6. Standard deviations obtained from three measurements of intervals (in milliseconds): between left edge and V-anchor, between C-center and V-anchor, and between right edge and V-anchor

	ti vs. *t-i*		
	speaker N.L.	speaker T.C.	speaker J.Y.
Left edge to V-anchor	9.15 < 11.25 (131) (136)	9.01 < 10.02 (105) (109)	8.81 < 10.69 (106) (103)
C-center to V-anchor	7.59 < 8.89 (108) (114)	8.66 < 10.71 (91) (93)	6.91 < 9.33 (91) (83)
Right edge to V-anchor	8.23 < 9.41 (85) (92)	9.97 < 12.34 (76) (76)	6.07 < 9.69 (77) (63)
	ni vs. *n-i*		
	speaker N.L.	speaker T.C.	speaker J.Y.
Left edge to V-anchor	7.81 < 9.21 (127) (137)	7.81 < 10.22 (107) (122)	12.06 > 9.92 (105) (96)
C-center to V-anchor	6.95 < 9.23 (109) (120)	6.67 < 8.62 (91) (103)	9.23 = 9.24 (89) (81)
Right edge to V-anchor	7.01 < 10.23 (91) (103)	7.43 < 9.28 (76) (85)	7.35 < 9.19 (73) (66)

Means are provided in parentheses (n = 14 for each test item).

any of the three measures. This suggests that the effect of the morphemic status on the timing variability holds for both *t* and *n*.

Amount of Contact in Pal- and i-Regions

In general, statistical analyses show that the degree of gestural overlap between tautomorphemic *t* and *i* is significantly less than the degree of overlap between their heteromorphemic counterparts. There was a significant main effect of morphemic status on both the Pal- and i-region contacts [for the Pal-region, $F_{(1, 52)} = 241.619$, $p < 0.0001$; for the i-region, $F_{(1, 52)} = 35.259$, $p < 0.0001$]. For the Pal-region, there was no effect of consonant identity, but there was a significant interaction between morphemic status and consonant identity [$F_{1, 52)} = 51.561$, $p < 0.0001$]; for the i-region, there was a significant effect of consonant identity [$F_{(1, 52)} = 16.315$, $p = 0.0002$] and a significant interaction between the two factors [$F_{(1, 52)} = 4.633$, $p = 0.0361$]. The significant interactions lead us to compare the consonants separately in order to examine how the effect of the morphemic status interacts with the consonant identity.

First, let us compare *ti* vs. *t-i* in the degree of gestural overlap. For *ti* vs. *t-i,* a one-way ANOVA shows significant differences between tautomorphemic and heteromorphemic sequences for both the Pal- and i-region contact amount [$F_{(1, 26)} = 744.737$, $p < 0.0001$ for the Pal-region; $F_{(1, 26)} = 22.388$, $p < 0.0001$ for the i-region]. Figure 12 shows that the percentage of contact is greater for heteromorphemic sequences than for tautomorphemic sequences, in both the Pal- and i-regions. This was true when each speaker was considered separately at $p < 0.0001$ for Pal-region contact. For i-region

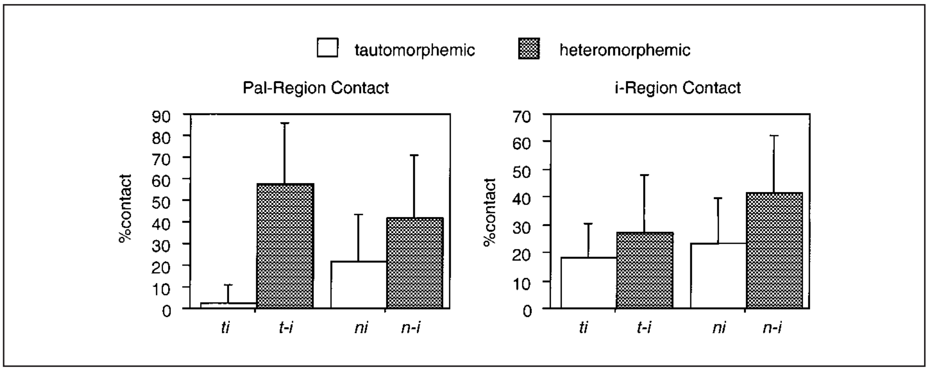

Fig. 12. Percent contact in the Pal-region (**a**) and the i-region (**b**), as a function of morphological structure. Data were pooled across speakers. (Note /t-i/ is realized as /t͡ʃi/ on the surface.) The error bars indicate standard deviations.

contact, 2 speakers (N.L. and T.C.), showed a significant difference at least at $p < 0.005$. For speaker J.Y., no comparison was made for i-region contact.[10] Overall, the patterns of contacts in both Pal- and i-regions indicate that the degree of gestural overlap between tautomorphemic *t* and *i* is less than the degree of overlap between heteromorphemic gestures. The difference between *ti* and *t-i* appears to be consistent with a phonological process of palatalization which has applied only to *t-i* sequence, resulting in [t͡ʃi].

What is more interesting is the *ni* vs. *n-i* contrast, which is also shown in figure 12. Contrary to the assumption in the previous literature (i.e. no morpheme boundary effect in a sequence of *n* and *i*), the degree of contact in both Pal- and i-regions is significantly less for the tautomorphemic sequence than for the heteromorphemic sequence [$F_{(1, 26)} = 21.811$, $p < 0.0001$ for Pal-region; $F_{(1, 26)} = 24.059$, $p < 0.0001$]. The significant difference remained when each speaker was considered separately at the level of $p < 0.005$.

So far we have observed significant contact differences in the Pal- and i-regions between tautomorphemic and heteromorphemic sequences (*ti* vs. *t-i* and *ni* vs. *n-i*), with less gestural overlap in tautomorphemic sequences. However, the degree of this effect appears to depend on the consonant involved, as indicated by a significant interaction between morphemic status and consonant identity [$F_{(1, 52)} = 4.633$, $p = 0.0361$]. Figure 12 shows that the amount of contact for both Pal- and i-regions is less for tautomorphemic *ti* than *ni*. Results of t tests showed that the contacts in both Pal- and i-region are significantly smaller for *ti* than for *ni* for speakers N.L. and J.Y. at least at $p < 0.05$. The same direction was also found for speaker T.C., but the difference was

[10] Her EPG data show that she produces *t-i* sequences by making a much broader central airway in order to produce frication, which presumably leads to no contact in the i-region. (Even at the midpoint of the vowel *i* in *t-i* context, the contact in the i-region was usually about less than 10%.) Thus, the comparison of contact amount in the i-region for within- vs. across-morpheme boundaries cannot be made for this speaker. Nevertheless, the very small amount in both Pal- and i-region contact (5 and 1%, respectively) in the tautomorphemic *ti* sequence for this speaker leads us to infer that only minimum gestural overlap has occurred.

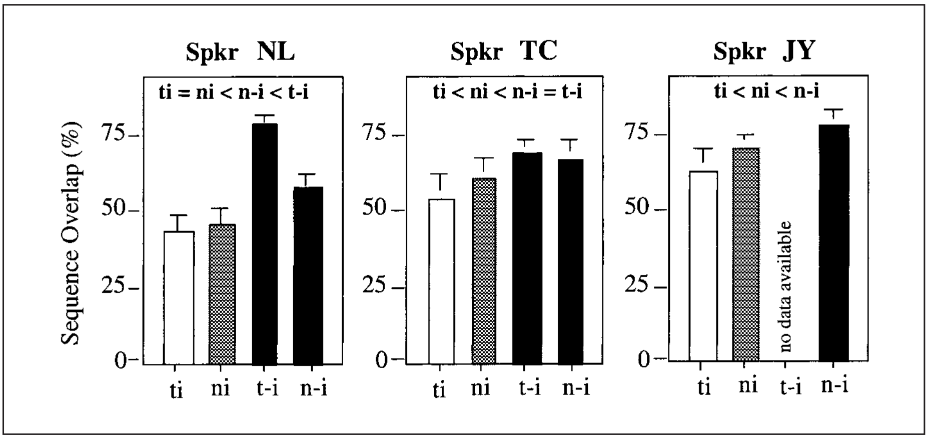

Fig. 13. Sequence overlap (%) during which contact occurs in both front region and i-region. < indicates statistical significance ($p < 0.01$) as obtained from unpaired t test and = indicates no significance. Data for *t-i* was not available for speaker J.Y. (Note that *t-i* is realized as [t͡ʃi] on the surface.) The error bars indicate standard deviations.

not statistically significant. This tendency suggests that the degree of gestural overlap varies depending on consonant identity, e.g. whether the target consonant is *t* or *n.*[11]

Sequence Overlap (%)

Results obtained from the measurement of sequence overlap (%) not only confirm the morpheme boundary effects but also show the different degrees of gestural overlap conditioned by the consonant type. The measurement of sequence overlap is not the measure of the amount of contact in a certain defined region, but rather the measure of temporal sequential overlap without consideration of the effects of the articulatory magnitude. The percent contact measure reflects the articulatory magnitude which may be influenced by the overall articulatory effort that the speaker made at the time he or she produces a sentence. Sequence overlap, on the other hand, reduces the possible influence of the overall articulatory effort because it considers only durational (temporal) overlap. A summary of the sequence overlap results is given in figure 13.

The first point to be made from figure 13 is that all 3 speakers distinguish a tautomorphemic sequence from a heteromorphemic sequence regardless of the target consonant. That is, the heteromorphemic *t-i* or *n-i* always has a greater sequence over-

[11] One may argue that the observed difference has nothing to do with gestural overlap but rather due to the intrinsic difference in place of articulation between *t* and *n,* i.e. *n* may be less anterior than *t.* However, the contact patterns for /pata/ and /pana/, which were produced as control tokens in isolation, do not show such difference at all for all 3 speakers. But one cannot exclude this possibility completely because we did observe more variable articulation in *n* than *t* elsewhere. In Cho and Keating [1999, to appear], in which the same 3 speakers participated, we examined the variation of consonant production as a function of prosodic position. We found that there is a shift in place, a backing as the stop moves from higher to lower domains. An informal observation confirmed that at least 1 speaker (N.L.) produced *n* in a more anterior direction than *t* in a lower domain (e.g. word-initially). However, no such difference was found in higher domains (e.g. intonation phrase-initially).

lap than their tautomorphemic counterparts (i.e. *ti*, *ni*). Second, speakers T.C. and J.Y. both display significant differences between *ti* and *ni*. Speaker N.L. does not make such a distinction. However, this speaker makes a clear distinction between *ti* and *ni* for both the Pal-region and i-region contacts, as shown in figure 12. Overall, when the results are pooled for all speakers, three categories can be drawn from the combination of the Pal- and i-region contacts and sequence overlap. The overlap is smallest for tautomorphemic *ti*, intermediate for tautomorphemic *ni*, and greatest for heteromorphemic *n-i* and *t-i*.

Summary of Experiment II

There are three main points that emerge from the results of experiment II. Firstly, the degree of variability in gestural timing in a *Ci* sequence is influenced by the hetero/tautomorphemic status of the sequence: the tautomorphemic sequence has less variability than the heteromorphemic one. This is in agreement with the findings of experiment I. Secondly, the degree of intergestural overlap also differs according to whether two adjacent gestures are tautomorphemic or heteromorphemic: the tautomorphemic sequence has less overlap than the heteromorphemic one.[12] Thirdly, the degree of intergestural overlap as revealed in the amount of contact in both i- and Pal-regions and the percent of sequence overlap is significantly different for *n* and *t:* the tautomorphemic *ti* sequence has less overlap than the tautomorphemic *ni* sequence.

General Discussion

It emerges from experiments I and II that intergestural timing is in general less variable between gestures *x* and *y* inside a single morpheme than across a morpheme boundary. Under the assumption that each morpheme is listed in the lexicon, the results suggest that gestural coordination is more *stable* within a single lexical entry than between lexical entries. Furthermore, there is a significant word-internal morpheme boundary effect on intergestural timing: that is, a homophonous word pair (e.g. *pan-i* vs. *pani*) shows different phonetic realizations sensitive to a morpheme boundary.

There are several significant implications made by this morpheme boundary effect. Firstly, it poses a serious problem on a traditional view [e.g. Chomsky and Halle, 1968; Kiparsky, 1982; Mohanan, 1982] which assumes that the morpheme boundaries are erased at the end of each transformational cycle, and that the relative timing of phonemes in the output is handled automatically by universally applicable phonetic module. A similar model has been proposed more recently by Levelt et al. [1999] in which the phonetic encoding does not have access to the morphological structure of the phonological form. As mentioned earlier, the model proposed by Levelt et al. [1999] retrieves a preprogrammed gestural score for the syllable from a reservoir, named 'syllabary', in which the morphological boundaries within a syllable are no

[12] In experiment I, in some cases the distance between C-center of /p/ and the end of the following vowel /i/ was greater across a morpheme boundary, which might suggest that there is *less* overlap for the heteromorphemic sequence. However, as noted earlier, the degree of overlap is a relative percentage in Articulatory Phonology. While the sequential overlap in experiment II is such a relative percentage measure, the absolute distance between gestures in experiment I does not necessarily provide an adequate measure of the degree of overlap. (In fact, the absolute distance between gestures for *ni* or *ti* in experiment II is also greater across a morpheme boundary.) Even if it turns out that there is an inconsistency between the /pi/ sequence and the /ni/ or /ti/ sequences, we can still claim that the stability of intergestural timing is the most important index of morphological structure, while the degree of overlap may vary with other factors (e.g. absence or presence of palatalization).

longer referred to. However, results of the present study suggest that the morphological structure is encoded in the phonetic details.

Secondly, in a similar vein, this study suggests that no type of morphemic concatenation is guaranteed to bring about neutralization (i.e. absence of phonetic difference) between tautomorphemic *xy* and heteromorphemic *x-y*. It has been claimed that no neutralization occurs between *xy* and *x-y* when the relevant boundary is a major phrase boundary or a word boundary [e.g. McLean, 1973; Hardcastle, 1985; Holst and Nolan, 1995; Byrd, 1998; Byrd et al., 1999], but this study demonstrates that the same may be true even for word-internal affixal boundaries. The evidence of such nonneutralization word-internally was found in variation in degree of *n*-palatalization depending on whether there is an affixal boundary or not.

Thirdly, we found that the degree of gestural overlap is greater for a *ni* sequence than for a *ti* sequence when both sequences are tautomorphemic (see footnote 11 for a possible alternative account). What are we to make of the fact that the degree of palatalization is different for /n/ than for /t/? Here I can suggest that such a difference comes from different output constraints: /n/ and /t/ have a different phonological status, because of an asymmetry in the phonemic inventory of Korean. To be more specific, the different degrees of overlap depending on the consonant identity fall out from two competing factors in the grammar: *paradigmatic contrast* vs. *preference for overlap*. Speakers have conflicting desires to make each gesture recoverable, and to compress words by overlapping gestures as much as possible. (One statement of the preference of gestural overlap may be found in Mattingly [1981], who suggests that overlap is preferred in order to maximize the parallel transmission of information; see Wright [1996] for a more extensive discussion.) On the other hand, the grammar also preserves some morphological information, so that it distinguishes *Ci* from *C-i* by allowing different degree of gestural overlap among them. In Korean, *t* is phonemically contrastive with t͡ʃ whereas *n* is not phonemically contrastive with ɲ. Thus, it is conceivable that the grammar keeps the contrast between *t* and t͡ʃ tautomorphemically by maximizing the difference in gestural overlap between them, preserving the paradigmatic contrast. By contrast, *n* does not have such a contrastive load, so that more overlap can be allowed in a tautomorphemic *ni* sequence. Thus, it is likely that the grammar allows greater gestural overlap for *n* than for *t*. (The fact that the place of articulation for *n* was sometimes found more variable than that for *t* also supports this idea; see footnote 11.) This idea is reminiscent of Lindblom's [1983] notion of *coarticulatory propensity*, according to which some segments are more likely to be coarticulated with neighbors than others, depending on the size and distribution of the phonemic inventory of a given language. Likewise, Manuel [1990] suggests that less coarticulation tends to be preferred in certain languages where extensive coarticulation would result in confusion of contrastive phonemes.

Finally, for those readers who are interested in the phonological implications of the asymmetry in *n*-palatalization, it should be noted that the present finding contradicts the prediction made by the traditional analysis, that *n*-palatalization is postlexical, and thus the degree of palatalization must be the same regardless of the morphological environment. However, *n*-palatalization may still be viewed as a 'postlexical' process if we adopt Zsiga's [1995] definition of the term 'postlexical'. Zsiga [1995] proposed that lexical palatalization as in 'confession' [kənfɛʃən] results from featural assimilation, a categorical process, and that postlexical palatalization as in 'confess you' [kənfɛʃu] is best accounted for by gestural overlap in a gradient fashion. In this connection, we can posit that the consistent pattern of *t*-palatalization in heteromorphemic *t-i* may be viewed as a lexical process, thus resulting in a categorical change from [t] to [t͡ʃ], whereas the asymmetry of *n*-palatalization between tauto- and heteromorphemic sequences may simply be because of the gradient aspect of gestural overlap depending on how close adjacent gestures are. While this cannot be tested systematically here, it would be reasonable to expect more overlap in *n*-palatalization for morpheme-internal gestures than for heteromorphemic gestures, since it is likely that the gestures of the former are more cohesive than gestures of the latter. However, the current study showed indeed the opposite pattern: the degree of *n*-palatalization is *smaller* morpheme-internally than across a morpheme boundary.

A possible explanation for this may be that a lesser degree of palatalization for morpheme-internal gestures has something to do with the stable organization among gestures inside a single lexical entry. Let us assume that such stable organization constrains the actual intergestural overlappings in such a way that substantial overlappings are blocked, whereas less stable organization among heteromorphemic gestures allows flexibility in gestural overlap, resulting in heteromorphemic palatalization. However, one problem is that flexibility in gestural overlap is not guaranteed to bring about greater gestural overlap. This might fall out from the principle of preference of overlap, as discussed in the preceding paragraphs. It is conceivable that speakers make as much gestural overlap as possible within a given limit, and that less constraint (or greater flexibility) on gestural overlap allows greater overlappings. This idea is consistent with a hypothesis made in Articulatory Phonology that gestural 'sliding' which results in 'perceptual' deletions of consonants or assimilations in casual speech [e.g. Browman and Goldstein, 1990] is most likely to occur at the end of one lexical unit and the beginning of the next one, a position in which gestures are coupled less strongly [Browman and Goldstein, 1998].

Models for Stability of Intergestural Timing

Thus far, I have discussed some implications of the current findings, but have set aside discussion of what mechanisms would account for the stability of intergestural timing associated with a single morpheme, or a lexical entry. A kernel from which the current paper has developed is the assumption made originally in Articulatory Phonology that lexical representations include specifications of temporal coordination among gestures. A working hypothesis throughout the paper is that gestures are coordinated more stably for a single lexical item (a morpheme or a lexicalized compound) than gestures spanning a boundary between lexical items, presumably because the phonological structure of a lexical entry forms a 'constellation' of gestures, a stable organization among gestures whose timing is specified in the lexicon. However, the exact mechanisms that relate lexically specified intergestural timing to their actual realization have not yet been fully spelled out. In what follows, I will discuss several competing, but potentially related hypotheses that may help us understand the stability of timing found for a single lexical entry.

Phase Window and Influencers

In Articulatory Phonology, the goal of the speaker's behavior is to make a 'particular organized ensemble of articulatory gestures' [Browman and Goldstein, 1992b,

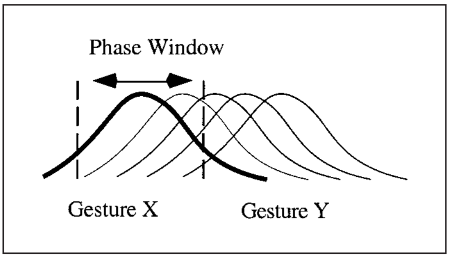

Fig. 14. A phase window allowing a range of overlap values for two gestures.

p. 222]. Gestures are coordinated with one another, and their specification includes both temporal and spatial information. Gestures can also overlap with each other, since they have internal durations. It is argued that many, if not all, phonological phenomena can be analyzed as changes in the magnitude or overlaps of gestures. For example, consonant assimilations can be viewed 'as a result of increasing gestural overlap between gestures on separate oral tiers' [e.g. Browman and Goldstein, 1992a, p. 361]. The mechanisms for coordinating gestures in Articulatory Phonology are *phasing rules*, which state that a phase in one gesture is synchronized with a phase in another gesture. However, Byrd [1996a] points out that rules specifying exact amounts of overlap are not sufficient, because the degree of overlap can vary depending on other linguistic and extralinguistic factors (e.g. syllable structure, stress, phrasing, speaking rate, etc.). For example, it has been found that the intergestural timing varies with syllable structure, i.e. the timing of gestures that constitute an onset consonant cluster is less variable than that of gestures forming a coda consonant cluster [e.g. Byrd, 1996b]. Byrd [1996b] suggests that this problem can be solved by adopting the notion of a *window* [Keating, 1990; Docherty, 1992; Byrd, 1994, 1996a]. Keating [1990] proposes that each segmental feature has a window, or a range of permissible spatial values, representing the contextual variability of a feature value. While her approach is primarily based upon spatial windows, temporal windows are defined in Docherty's and Byrd's approaches. In the *Phase Window model* developed by Byrd [1994, 1996a] for English consonant sequences, one of the sources of variability in gestural overlap may be a wide range of possible timing relationships, i.e. a wide phase window. A hypothetical phase window, as indicated by an arrow, is schematized in figure 14.

Byrd [1996a] argues that a relatively narrow window can be due to the lexically specified timing, and therefore allows less variability. However, in her Phase Window model, such a lexically specified narrow phase is limited to gestures which form 'what have been traditionally considered to be a segment' (e.g. a complex segment /$\widehat{\text{mp}}$/ or a simple segment /m/ or /p/ with multiple gestures).

In addition to the size of the phase window, another source of the variability comes from a variety of linguistic and extralinguistic factors (e.g. syllable structure, phrase boundary, stress, speaking rates, etc.) or *influencers*, that weight the phase window differently. This weighting determines where in the range of possible values the actual intergestural overlap will be implemented, and how narrow the range will be. In short, the variability in intergestural timing is influenced by two independent factors: (a) the window width (the narrower, the less variable) and (b) the weighting of the influencers (the more narrowly the region within the phase window is weighted, the less variable).

The actual realization of the intergestural timing, however, is complex, and is not yet spelled out clearly, as noted by Byrd [1996a, p. 151], who states '[o]f course, an interesting empirical and theoretical question is how this determination of the combined weighting of the phase window is arrived at'. The outline of how the actual gestural timing may be related to the phase window and the influencers follows. First, the left and right limits of the phase window for a certain temporal coordination of two adjacent gestures is learned by the child who acquires language-specific constraints imposed on the coordination of gestures involved. Then, the combined effect of the influencers existing for the particular gestural sequence determines the probability density), i.e. where in the window phasing occurs most probably. Byrd [1996a, p. 151] explains:

> The more alike the contextual effects are from token to token, the more alike the combined influencer distributions will be. This will yield a high probability of similar organizations being realized in similar contexts – i.e. low token-to-token variability.

Now let us consider how more stable timing within a single lexical entry than across different lexical entries, as found in this study, can be achieved under the current Phase Window model. The only allowable way is through the weighting of a narrower region within a fixed phase window. (Recall that Byrd proposes that the phase window is lexically specified narrowly only for the gestures associated with a single segment.) It is then possible that the presence or absence of a morpheme boundary acts as a weighting factor (though the effect of morpheme boundary is not included explicitly as one of known or potential influencers in Byrd's model). Specifically, there would be a narrower region weighted for adjacent gestures inside a morpheme than for those across a morpheme boundary, while the size of the phase window would remain fixed. This is similar to what Byrd [1996a] predicts about the timing of the gestures within words compared to that of the same gestures between words.

However, as noted earlier, Byrd's [1996a] model cannot be tested until we know exact influencers and mechanisms of the weighting function that narrows or widens the region. It might in fact be impractical to discover what all the influencers are and how they interact with each other to generate the final density of probability distributions. To the extent that the theory holds up, there are still unclear aspects under Byrd's model. Firstly, it is not clear why the lexically specified narrow phase window is limited to the phasing of gestures that constitute a segment, and why it cannot be extended to the phasing of gestures inside a lexical item. (But see below for Zsiga's potential phonological explanation for this.) There are two mechanisms that derive the stability of the timing: one through the window width and another through weighting factors. A simpler model may be the one that invokes only one of the mechanisms. Relatedly, seen from a slightly different angle, it is unclear why the stability for a segment cannot be derived by another weighting factor, say, the 'segmenthood' of gestures, rather than resorting to narrowing the window width. If the intergestural timing between segments inside a single lexical item was determined by weighting function, there would also be no theory-internal reason to prevent the region from being more narrowly weighted for the gestures constituting a segment than for the same gestures across segments. Of course, this is in direct contradiction with Byrd [1996a, p. 160], who proposes that 'it is not the case that the quality of being a segment causes stable timing, but rather that stable timing causes the quality of being a segment'. However, the rationale behind this claim is not explicitly spelled out. How do we decide that timing is stable enough to cause the quality of being a segment? To support this idea, Byrd states, citing Saltzman and Munhall [1989], that the cohesion of gestures for segmental units is due to dynamical coupling of the gestures, and that the integrity of gestures forming a segment is maintained in fluent speech. The crucial assumption is that 'an appropriate type of coupling could yield a limited window of relative phase relations between the coupled gestures' [Byrd, 1996a, p. 160]. But it remains still unclear how 'appropriate' the type of coupling should be in order to enter into the lexicon.

Extended Lexically Specified Phase Window

An alternative is for the model to allow a narrower phase window for gestures inside a lexical item (a morpheme or a lexicalized compound), but a wider phase window for those across lexical items. One way of achieving this is through the assumption in Articulatory Phonology that intergestural timing is specified in the lexicon. Unlike Byrd's Phase Window model that states that a narrow window due to lexically specified gestural timing is limited to gestures constituting a segment, this view assumes that such a narrow phase window is extended to gestures constituting a lexical item, i.e. the phase window is specified more narrowly within a lexical entry than across lexical entries. Then, greater stability can be understood as a result of a narrower phase window inside the lexicon than outside the lexicon, all else being equal. This alternative view also agrees with Ladefoged's [1992, p. 174], view that timing between two adjacent segments belonging to the same lexical entry (or constituting a word) can be taken out of the lexicon as 'pre-programmed chunks that can be triggered off as a whole'. It is conceivable that language learners are exposed to the stable timing of gestures within a lexical entry, and lexicalize such a stable intergestural timing. However, what is stored in the lexicon is not the specific timing relationship (perhaps through phasing rules) but an empirically determined permissible range of intergestural timing. All else being equal, the difference in the size of the phase window would be a major determinant of intergestural timing variability.

There arise a few questions, however, in this view. The first question is about stability associated with syllable structure. It has been reported that timing relations between syllable onset consonant clusters are more stable than those between syllable coda consonant clusters [e.g. Byrd, 1996a, b]; and that the onset consonant-to-vowel (CV) relations tend to be more stable than the vowel-to-coda consonant (VC) relations [e.g. Browman and Goldstein, 1988; Byrd, 1995]. The question is then how the systematic timing patterns arising from syllable structures can be accounted for if intergestural timings are explicitly specified for all relations in each and every lexical entry. Byrd [pers. commun.] has pointed out that if the onset CV relation, for example, always shows the same type of timing arrangement, this would just be a coincidence if every CV relationship were specified independently in every lexical entry. If this explanation were correct, the idea of moving the specification of timing relations into the lexical entry would require further ad-hoc mechanisms to capture generalizations concerning syllable structure. However, the results of the present study suggest that the CV relations do not always show the same type of timing patterns: the CV timing is more stable within a single lexical entry than across different lexical entries, found in both EMA and EPG data. It is possible that, as Byrd predicts, in deriving the different CV timing patterns, there still would be a narrower region weighted for the CV phasing than for the VC phasing, resulting in more systematically stable CV timings. Nonetheless, if such weighting for the CV phasing takes place within a narrow phase window for a lexical item, but within a relatively wider phase window for gestures across different lexical items, all else being equal, we may still account for the observed timing differences found in the present study.

Another fundamental question is whether less variability in timing is sufficient grounds for claiming that the timing relation must be specified in the lexicon in one case (e.g. within a morpheme) and not in the other (e.g. across a morpheme boundary). The same question can be addressed to the Byrd's Phase Window model which argues that the intergestural timing is lexically specified only for a segment. Certainly, we do not have empirical evidence for that, and much work needs to be done, including developing the mechanism that determines how stable the timing needs to be in order to enter the lexicon, as was the case with the Phase Window model.

Bonding Strength

There is yet another way of handling variability in intergestural timing, proposed by Browman and Goldstein [1998]. In Articulatory Phonology, the specification of temporal coordination occurs locally in a pairwise fashion: '[g]iven this fact, for an utterance with *n* gestures, specification of *n-i-1* pairs completely determines the temporal structure of an utterance' [Browman and Goldstein, 1998, p. 2]. They propose that the variability that may occur across pairs of phasing relations due to some other linguistic or extralinguistic factors can be handled by associating every phase relation within a lexical entry with a 'bonding strength', or the degree of cohesion of gestures involved. As in Byrd's Phase Window model, a given pair of gestures is hypothesized to be influenced, or weighted, by the sources of timing variability (e.g. speaking rate, style, prosody) in inverse proportion to their bonding strengths. They further explain that the lesser variability in timing for onset consonant gestures than for coda consonant gestures is due to greater bounding strength of onset consonant gestures. For example, if greater variability among coda consonant gestures than among onset consonant gestures would come from local speaking rate changes, then such a speaking rate effect would be greater on coda consonant gestures due to their weaker bonding strength. This bonding strength analysis is argued to account for the environments in which gestures could 'slide' in casual speech, resulting in perceptual 'hiding' of consonant gestures and assimilations [Browman and Goldstein, 1990, 1998, p. 3]. They state that:

> In fact, in all these cases [of the perception of consonant deletion and assimilations], the gestures that slide with respect to one another are not part of the same lexical unit. If we hypothesize that post-lexical phasing between gestures at the end of one lexical unit and beginning the next has weak (or possibly non-existent) bonding strength, then the site of these assimilations and deletions is accounted for.

What is relevant to the current study is that postlexical phasing among gestures across different lexical entries has *weaker* bonding strength than lexical phasing among gestures within a single lexical entry. The stronger bonding for gestures within a lexical entry then would result in a greater stability in intergestural timing as compared to the weak bonding for gestures across different lexical entries. If such bonding strength is added to phasing rules, timing relations can be made more flexible, allowing variability without having recourse to the Phase Window.

Evidence which lends support to this hypothesis may be found in the fact that intervals between gestures tend to be smaller within a word or a morpheme than across word or morpheme boundaries. The shorter intervals associated with a single lexical entry appear to suggest that gestures are more strongly bonded within a lexical entry. This bonding strength hypothesis also opens up the possibility that the timing relationships may be constrained by physical factors, rather than control factors. If strongly bonded gestures are necessarily associated with the short gestural intervals, the more stable timing between gestures may simply be a by-product of such short intervals. However, there are no explicit mechanisms yet available that relate the bonding strength. In addition, the model needs to allow the morphological structure as a potential factor that influences bonding strength. To the extent that it holds up, the bonding strength hypothesis seems to be compatible with the findings of the current study.

Thus far, we have outlined how stability of timing may be represented using phase windows or possibly bonding strength. We suggested, as one of many other possibilities, that the stability in gestural timing may be due to specification of gestural timing in the lexicon. At this point, it is worth discussing a related problem regarding an unusual prediction made by the idea of lexically specified timing. That is, if the timing relationships among gestures of a lexical entry were specified in the lexicon, the same string of gestures forming different lexical entries, or homophones, could have different intergestural timing. Although this may sound bizarre, the present findings at least suggest that it may not be entirely infeasible that such intergestural timing is specified in the lexicon.[13] Evidence could be found elsewhere, too, e.g. in Guion [1995], who showed that timing may vary between homophonous words (e.g. *need* vs. *knead, way* vs. *whey,* see below).

Relatedly, if a particular pattern characterizes an entire class of morphemes in the language, it is generally not analyzed as a lexical idiosyncrasy (this was pointed out by a reviewer). Traditionally [e.g. Chomsky and Halle, 1968], lexically specified properties are idiosyncratic prosodic and featural attributes that are potentially contrastive and cannot be predicted from other properties of the string. Thus, if the stability of timing is characteristic of all monomorphemic words, it is conceivable that such a systematic pattern is predictable and thus grammaticized in the language, rather than specified separately for each lexical item in the lexicon. Zsiga's [1997] phonological timing model, outlined below, is in fact compatible with this view.

Phonological Timing as a Source of Gestural Tightness

Zsiga [1997, p. 268] suggests that 'phasing among articulators works outward from the smallest domain to the largest'. It is assumed that the degree of cohesion among gestures forming a segment will be greatest, and, therefore, result in the most stable intergestural timing, and that the tightness decreases as a prosodic domain moves up in the hierarchy (e.g. segment, syllable, foot, word, phrase, etc.). Unlike the idea of lexically specified timing, what is crucial in Zsiga's model is that the source of gestural

[13] Other, rather indirect, evidence supporting the specification of noncontrastive properties such as timing in the lexicon comes from Fougeron and Steriade [1997], who examined phonetic details (e.g. interconsonantal timing and linguopalatal contact) of C′C sequences (e.g. *pas* ***d'rôle***, where the apostrophe marks the position of a deleted schwa) created by schwa deletion and other CC sequences (both underlying CC and C#C, where # is a word boundary). The results show that some durational, articulatory and/or timing properties of the full form (e.g. *pas* ***d**ə* ***r**ôle* 'no role') are maintained in the form with schwa deletion (e.g. *pas* ***d'r**ôle*) which in turn show phonetic properties distinct from those of not only underlying the CC cluster (e.g. *pas* ***dr**ôle* 'not funny') but also the C#C cluster. Fougeron and Steriade [1997, p. 946] conclude that:

> This invariant character of the duration and spatial magnitude of the [d] closure in [de]/[d′] suggests to us that these properties are present in the lexical entry of [de], despite their non-contrastive character. Thus we conjecture that non-contrastive properties – properties which by themselves cannot support a phonemic contrast – may nonetheless be present in lexical entries.

tightness (or stability) comes from categorical timing relations in phonology which are indicated by the presence or absence of association lines. For example, the tightness of gestures that forms a segment is due to a single root node associated with features that are later translated into gestures on the phonetic level. (Note that while Zsiga uses association of features to a single node as an explanation for the most stable timing of gestures forming a segment, Byrd [1996a] employs a lexically specified phase window to account for the stable timing.)

In this framework, stability within a lexical entry would come from more tightness among gestures inside a word than across a word, if phonological words are the only type of lexical entry. However, Zsiga's model excludes a potential morpheme boundary effect, since the morpheme boundary inside a syllable should be erased if syllables are coordinated with each other. To the extent that the morpheme boundary effect is incorporated into the model, the variability of the gestural tightness could account for the stability found for a single lexical entry. Crucially, the source of the stability would still come from a categorical timing relationship in phonology, rather than from lexical specification of intergestural timing.

However, as also noted by Zsiga, the exact mechanisms of how root nodes map gestural units and how the tightness is related to the actualization of the phasing remains still unclear, which renders her model untestable at the moment.

Frequency of Occurrence and Variability

Quite independently, another reasonable source of the stability of intergestural timing may come from how familiar the speakers are with items that are tested. It has been well known that speakers may produce more frequently used words with less extreme articulation or greater reduction than when they produce less frequently used words [e.g. Balota et al., 1989; Guion, 1995 and others cited therein; Wright, 1997]. Unfamiliar words are often produced with extreme articulation which accompanies lengthening, a kind of hyperarticulation [cf. Lindblom, 1990], and such an effect occurs even when test words are homophones with different frequencies of occurrences [Guion, 1995]. This seems to be consistent with the small but in some cases significant lengthening associated with the nonlexicalized compounds found in experiment I. In this connection, the timing difference in experiment I might have come about simply because the lexicalized compounds tested were familiar to the speakers whereas the nonlexicalized ones were not. It may well be the case that speakers produce lexicalized compounds with less variability than nonlexicalized compounds because they produced the former more frequently (i.e. practice effect).

This frequency effect, however, may be compatible, or interrelated, with other hypotheses discussed earlier. Let us first consider the Phase Window model. Recall that probability density (i.e. where in the window phasing occurs most probably) is determined by the combined effect of the influencers. It is then plausible that frequently heard lexical items (e.g. lexicalized compounds) provide a high probability of similar organizations, which yields low token-to-token variability. Thus, frequency of co-occurrence of adjacent gestures may act as an influencer. Secondly, the frequency effect appears to be related to the extended lexically specified phase window. It is assumed that the lexicalization of the phase window is done through exposure to the timing of gestures in the course of language acquisition. It is conceivable that children are exposed to lexicalized compounds frequently, and eventually acquire a relatively narrower phase window for frequently heard lexical items. Similarly, it may well be the case that bonding strength may increase among gestures inside high frequency words, whereas it may decrease among gestures of low frequency words.

Thus far, we have considered several hypotheses which might account for the stability of the timing of gestures as a function of lexical and/or morphological status of units in which gestures occur. Unfortunately, all these hypotheses are currently untestable because we do not have enough information of how each model assumed works, i.e. methodology to determine relevant influencing factors (in the case of Phase Window model and Bonding Strength model) and the interaction between such factors and the window width (in the case of Extended Lexically Specified Phase Window model). While the debate over what could be the best account for the stable timing within a single lexical item cannot be resolved here, the discussion provided in preceding paragraphs should make predictions that would hopefully inspire future research in intergestural timing.

Conclusion

The current study has examined the effect of morpheme and word boundaries on the variability of intergestural timing. The results of two experiments show that there is a significant effect of morpheme boundaries on intergestural timing, i.e. more stable timing within a single morpheme (which is viewed as a single lexical entry) than across a morpheme boundary. Furthermore, the phonetic details of gestural overlap shed light on the asymmetry of palatalization between *ti* and *ni* in Korean. In particular, the effects of affixal morpheme boundaries on palatalization patterns for *ni* show that no neutralization is guaranteed within a phonological word.

Overall, the current study implies that the morphological structure may be encoded in the phonetic realization. This may seem radically different from traditional phonological theory [such as Chomsky and Halle, 1968], which assumes that the relative timing of phonemes in the output is handled automatically by the phonetic module, and therefore there are no effects of morpheme boundaries on it. Nevertheless, the findings reported in this paper open up the possibility that the phonetic realization is indeed influenced by the morphological structure. It is hoped that the current study will inspire future research that explores further the potential encoding of the morphological structure in low-level phonetics across languages.

Acknowledgments

I would like to thank our subjects Hyejeong Han, Hyuckjoon Lee, Namhee Lee, and Jiyoung Yoon for participating in the experiments. Special thanks go to Adam Albright for pointing out every detailed matter throughout the paper. This paper has also greatly benefited from discussion with and comments from Dani Byrd, Bruce Hayes, Sun-Ah Jun, Pat Keating, Donca Steriade. This paper has also been strengthened by thorough and insightful comments from the editor, John Kingston, reviewer, Caroline Smith, and two other anonymous reviewers. I wish to thank Marco Baroni, Melissa Tai and Siri Tuttle for proofreading and valuable suggestions, and other members of the UCLA Phonetics Lab for comments on earlier presentations of the work. This work was supported by the Department of Linguistics, UCLA.

References

Ahn, S.-C.: The interplay of phonology and morphology in Korean; PhD diss. University of Illinois at Urbana-Champaign (1985).

Balota, D.A.; Boland, J.E.; Shields, L.W.: Priming in pronunciation: beyond pattern recognition and onset latency. J. Memory Lang. *28:* 14–36 (1989).

Barry, M.: Palatalization, assimilation and gestural weakening in connected speech. Speech Commun. *11:* 393–400 (1992).

Bhat, D.N.S.: A generalization of palatalization; in Greenberg, Ferguson, Moravcsik, Universals of human language, vol. 2, pp. 47–92 (Stanford University Press, Stanford 1978).

Booij, G.E.: Word formation. Lingua *65:* 260–270 (1985).

Browman, C.; Goldstein, L.: Toward an articulatory phonology. Phonol. Yb. *3:* 219–252 (1986).

Browman, C.; Goldstein, L.: Some notes on syllable structure in articulatory phonology. Phonetica *45:* 140–155 (1988).

Browman, C.; Goldstein, L.: Articulatory gestures as phonological units. Phonology *6:* 201–252 (1989).

Browman, C.; Goldstein, L.: Tiers in articulatory phonology, with some implications for casual speech; in Kingston, Beckman, Papers in Laboratory Phonology. I: Between the grammar and the physics of speech, pp. 341–376 (Cambridge University Press, Cambridge 1990).

Browman, C.; Goldstein, L.: Articulatory phonology: an overview. Phonetica *49:* 155–180 (1992a).

Browman, C.; Goldstein, L.: Responses to commentaries. Phonetica *49:* 222–234 (1992b).

Browman, C.; Goldstein, L.: Dynamics and articulatory phonology; in Port, van Gelder, Mind as motion: explorations in the dynamics of cognition, pp. 175–193 (MIT Press, Cambridge 1995).

Browman, C.; Goldstein, L.: Competing constraints on intergestural coordination and self-organization of phonological structures. Current Trends in Phonology, Royaumount, 1998. (To appear in Bull. Commun. parlé.)

Byrd, D.: Articulatory timing in English consonant sequences; PhD diss. UCLA (1994). UCLA Working Papers in Phonet. *86* (1994).

Byrd, D.: C-Centers revisited. Phonetica *52:* 285–306 (1995).

Byrd, D.: A phase window framework for articulatory timing. Phonology *13:* 139–169 (1996a).

Byrd, D.: Influences on articulatory timing in consonant sequences. J. Phonet. *24:* 209–244 (1996b).

Byrd, D.: Intergestural coordination adjacent to multiple boundaries. Proc. 16th Int. Congr. Acoustics and 135th Meeting Acoust. Soc. Am., vol. 2, pp. 1255–1256 (1998).

Byrd, D.: Articulatory vowel lengthening and coordination at phrasal junctures. Phonetica *57:* 3–16 (2000).

Byrd, D.; Kaun, A.; Narayanan, S.; Saltzman, E.: Phrasal signatures in articulation; in Broe, Pierrehumbert, Papers in Laboratory Phonology V: Language acquisition and the lexicon, pp. 70–87 (Cambridge University Press, Cambridge 2000).

Byrd, D.; Saltzman, E.: Intragestural dynamics of multiple prosodic boundaries. J. Phonet. *26:* 173–199 (1998).

Cho, T.: Effect of prosody on vowel-to-vowel coarticulation in English. Proc. Int. Congr. Phonet. Sci. (ICPhS) 99, vol. 1, pp. 459–462 (1999).

Cho, T.; Keating, P.: Articulatory and acoustic studies of domain-initial strengthening in Korean. UCLA Working Papers in Phonet. *97:* 100–138 (1999). (To appear in J. Phonet. 29.)

Chomsky, N.; Halle, M.: The sound pattern of English (Harper & Row, New York 1968).

Dixon, W.J.: BMPD statistical software manual (The Regents of University of California, Los Angeles 1988).

Docherty, G.J.: The timing of British English obstruents (Forsi, Berlin 1992).

Edwards, J.; Beckman, M.E.; Fletcher, J.: The articulatory kinematics of final lengthening. J. acoust. Soc. Am. *89:* 369–382 (1991).

Flege, J.E.: Effects of speaking rate on tongue position and velocity of movement in vowel production. J. acoust. Soc. Am. *84:* 901–916 (1988).

Fougeron, C.; Steriade, D.: Does deletion of French schwa lead to neutralization of lexical distinction? In Proc. Eurospeech 97, pp. 943–946, Rhodes 1997.

Gay, T.; Ushijima, T.; Hirose, H.; Cooper, F.S.: Effect of speaking rate on labial consonant-vowel articulation. J. Phonet. *2:* 47–63 (1974).

Guion, S.G.: Word frequency effect among homonyms. Tex. Linguistic Forum *35:* 103–115 (1995).

Hardcastle, W.J.: Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. Speech Commun. *4:* 247–263 (1985).

Holst, T.; Nolan, F.: The influence of syntactic structure on [s] and [ʃ] assimilation; in Connell, Arvaniti, Papers in Laboratory Phonology. IV: Phonology and phonetic evidence, pp. 315–333 (Cambridge University Press, Cambridge 1995).

Iverson, G.; Wheeler, D.W.: Blocking and the elsewhere condition; in Hammond, Noonan, Theoretical morphology, pp. 325–338 (Academic Press, San Diego 1988).

Jun, S.-A.: Asymmetrical prosodic effects on the laryngeal gesture in Korean; in Connell, Arvaniti, Papers in Laboratory Phonology. IV: Phonology and phonetic evidence, pp. 235–253 (Cambridge University Press, Cambridge 1995).

Keating, P.A.: The window model of coarticulation: articulatory evidence; in Kingston, Beckman, Papers in Laboratory Phonology. I: Between the grammar and the physics of speech, pp. 451–470 (Cambridge University Press, Cambridge 1990).

Keating, P.A.: Coronal places of articulation; in Paradis, Prunet, Phonetics and phonology. II. The special status of coronals: internal and external evidence, pp. 29–48 (Academic Press, San Diego 1991).
Keating, P.A.; Cho, T.; Fougeron, C.; Hsu, C.: Domain-initial strengthening in four languages. UCLA Working Papers in Phonet. *97:* 139–151 (1999). (To appear in Papers in Laboratory Phonology VI.)
Kim, S.-G.: Palatalization in Korean; PhD diss. University of Texas, Austin (1976).
Kim-Renaud, Y.-K.: Korean consonantal phonology; PhD diss. University of Hawaii (1974).
Kiparsky, P.: Elsewhere in phonology; in Anderson, Kiparsky, A Festschrift for Morris Halle, pp. 93–106 (Holt, Rinehart and Winston, New York 1973).
Kiparsky, P.: Lexical phonology and morphology; in Yang, Linguistics in the morning calm, vol. I, pp. 3–91 (Hanshin, Seoul 1982).
Kiparsky, P.: Blocking in non-derived environment; in Hargus, Kaisse, Studies in lexical phonology, pp. 277–314 (Academic Press, San Diego 1993).
Ladefoged, P.: The main interfaces between phonetics and phonology; in Dressler, Luschutzky, Pfeiffer, Rennisson, Phonologica 1988, pp. 165–179 (Cambridge University Press, Cambridge 1992).
Lahiri, A.; Evers, V.: Palatalization and coronality; in Paradis, Prunet, Phonetics and phonology. II. The special status of coronals: internal and external evidence, pp. 79–100 (Academic Press, San Diego 1991).
Levelt, W.J.M.; Roelofs, A.; Meyer, A.S.: A theory of lexical access in speech production. Behav. brain Sci. *22:* 1–75 (1999).
Levene, H.: Robust tests for equality of variance; in Olkin, Contributions to probability and statistics (Stanford University Press, Palo Alto 1960).
Lindblom, B.: Economy of speech gestures; in MacNeilage, The production and modeling of speech, pp. 217–245 (Springer, New York 1983).
Lindblom, B.: Explaining phonetic variation: a sketch of the H&H theory; in Hardcastle, Marchal, Speech production and speech modeling, pp. 403–439 (Kluwer Academic Publishers, Dordrecht 1990).
McLean, M.: Forward coarticulation of velar movement at marked junctural boundaries. J. Speech Hear. Res. *16:* 286–296 (1973).
Manuel, S.: The role of contrast in limiting vowel-to-vowel coarticulation in different languages. J. acoust. Soc. Am. *88:* 1286–1298 (1990).
Mattingly, I.G.: Phonetic representation and speech synthesis by rule; in Myers, Laver, Anderson, The cognitive representation of speech, pp. 415–420 (1981).
Miller, J.L.: Effects of speaking rate on segmental distinctions; in Eimas, Miller, Perspectives on the study of speech, pp. 39–74 (Erlbaum, Hillsdale 1981).
Mohanan, K.P.: Lexical phonology; PhD diss. MIT (1982).
Perkell, J.S.; Cohen, M.H.; Svirsky, M.A.; Matthies, M.L.; Garabieta, I.; Jackson, M.T.T.: Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. J. acoust. Soc. Am. *92:* 3078–3096 (1992).
Pierrehumbert, J.; Talkin, D.: Lenition of /h/ and glottal stop; in Docherty, Ladd, Papers in Laboratory Phonology. II: Gesture, segment, prosody, pp. 90–117 (Cambridge University Press, Cambridge 1992).
SAS Institute Inc.: StatView 5.0 SAS Institute Inc. (Cary 1998).
Saltzman, E.; Munhall, K.: A dynamical approach to gestural patterning in speech production. Eco. Psychol. *1:* 333–382 (1989).
Schoenle, P.W.: Elektromagnetische Artikulographie (Springer, Berlin 1988).
Schoenle, P.W.; Mueller, C.; Wenig, P.: Echtzeitanalyse von orofacialen Bewegungen mit Hilfe der elektromagnetischen Artikulographie. Biomed. Technik *34:* 126–130 (1989).
Tuller, B.; Shao, S.; Kelso, J.A.S.: An evaluation of an alternating magnetic field device for monitoring tongue movements. J. acoust. Soc. Am. *88:* 674–679 (1990).
Wright, R.: Consonant clusters and cue preservation in Tsou; PhD diss. University of California, Los Angeles (1996).
Wright, R.: Lexical competition and reduction in speech: a preliminary report. Research on Spoken Lang. Processing, Progr. Rep. 21, pp. 471–485 (Indiana University, 1997).
Zsiga, E.: Features, gestures, and the temporal aspects of phonological organization; PhD diss. Yale University (1993).
Zsiga, E.: An acoustic and electropalatographic study of lexical and postlexical palatalization in English; in Connell, Arvaniti, Papers in Laboratory Phonology. IV: Phonology and phonetic evidence, pp. 282–302 (Cambridge University Press, Cambridge 1995).
Zsiga, E.: Features, gestures, and Igbo vowels: an approach to the phonology-phonetics interface. Language *73:* 227–274 (1997).