# Revisiting Distinctive Feature Approach in Speech Recognition*

Kee-Ho Kim

This paper intends to show how current phonological theories of feature hierarchy and underspecification can be implemented into speech recognition so that we can take advantage of both invariant and allophonic cues for parsing and matching. I first show the usefulness of variant cues in parsing without relying on the higher level constraints, and I show that the distinctive feature approaches are better than the segmental approaches in confusion error analyses as well as in utilizing phonological generalizations. Mainly due to the defects of the feature theory *per se*, the feature analysis has been neglected in speech recognition, and thus I argue that the binary notion of feature values and the requirement of fully specified phonetic representation should be given up, and instead the privativeness notion of features be adopted, which leads to surface phonetic underspecification.

## I. Introduction

It is generally agreed that there are two types of phonetic (and phonological) cues in speech recognition: a) the invariant cues which are relatively invariant to context (e.g. place, manner), and b) the variant (or allophonic) ones which vary a great deal with context (e.g. aspiration, flapping in English; voicing in Korean).

Many phonetic and phonological works show that phonemes generally have different acoustic realizations depending on the context. For example, the phoneme /t/ in 'top' in Englsih is typically realized as heavily aspirated syllable-initially, but it is realized without any aspiration burst in syllable-final position such as in a word like 'pot'. The phoneme /t/ in /tap/ 'answer' in Korean is realized with slight aspiration in word-initial position, while it is realized as a

* This paper has been presented at SICONLP '90.

voiced one between sonorant segments as in /potap/ [podap] 'recompense'. In early 70s (during the ARPA speech project), however, these kinds of variations are considered to be problematic for speech recognitions, and hence these allophonic cues are regarded as a kind of 'noise'. (Klatt (1980a, b), Cole and Jakimik (1980); for example, in Stevens' (1981) theory of invariant features, only invariant features are assumed to be distinctive and thus useful for retrieving the word from the lexicon, while variant features are assumed to be non-distinctive and thus superfluous for lexical retrieval.

Recent works (Nakatani and his co-workers (1977, 1978); Church 1987), however, show that variant, or allophonic cues are also useful in speech recognition just like invariant cues; Church (1987) argues that variant features such as aspiration and flapping in English are useful for parsing the utterance into syllables and metrical feet, and invariant features such as place and manner are useful for matching against the lexicon.

In what follows, I will first recapitulate some of Church's recent arguments for the usefulness of the allophonic cues, and show how the allophonic cue of voicing can be effectively used for parsing in Korean. And in section 3, still using some of his arguments, I will show that the distinctive feature approaches are better than the segmental approaches in confusion error analyses as well as in capitalizing phonological generalizations. Then in section 4, I argue that, in order to avoid certain defects of the distinctive feature theory (SPE, Chomsky & Halle (1968)), the binary notion of feature values and the requirement of fully specified phonetic representation should be given up, and instead the privativeness notion of features be adopted. Finally, in section 5, I will discuss both phonological and surface phonetic underspecification.

## II. Allophonic Cues are Useful in Speech Recognition

As shown before, phonemes might have several allophones, or contextual variants; i.e. they can have different phonetic realizations depending on the context. For example, the phoneme /t/ as in *top* has strong aspiration while the one in *stop* and *pot* does not have. However, in English, there are no words distinguished just on the basis of aspiration with regard to stop consonants. In other words, there are no minimal pairs with the contrast of only aspiration. So aspiration is not distinctive in English and thus it is regarded as a contextual variant, or allophonic one. Contrary to English, the feature aspiration in Korean is distinctive, since /p$^h$ul/ '풀 (grass)' and /pul/ '불 (fire)' are minimal pairs, which have all features in common except the feature aspiration. What we note here is that the same phonetic feature aspiration may be regarded as either a variant cue (as in English) or an invariant one (as in Korean), depending on the languages. Similarly, the feature voicing may be regarded as either a variant cue or an invariant one. That is, in English, voicing is distinctive and thus regarded as an invariant cue since /tip/ *tip* and /dip/ *dip* are minimal pairs. But, in Korean, it is regarded as a variant cue since it is not distinctive, or redundant. Of interest here is that, regardless of whether they are variant or invariant cues, these manner features are very easy to detect in spectrogram analyses. That is, aspiration and voicing features have their own particular acoustic cues: aspiration with quite long duration of 'noise' after the release of the closure (long positive VOT), and voicing with little bars in the very low frequencies.

Nevertheless, in early 70s, this kind of allophonic cues had been regarded simply as a kind of 'noise', useless for the lexical retrieval. Therefore, in order to disambiguate pairs such as *at ease* vs. *a tease*, *night rate vs. nitrate*, *sheep raid* vs. *she prayed* vs. *sheep preyed*, etc., heavy relying on higher level constraints such as syntactic and semantic ones was required. Recently, however, Nakatani and his co-workers (1977, 1978) and Church (1987) have shown that these allophonic cues are very useful in parsing the utterances into syllables without any relying on syntactic and semantic constraints.

Consider the following English examples in (1), where each pair has phonemically identical sequences of segments.

| | | |
|---|---|---|
| (1) | sheep raid vs. she prayed | /ʃipreid/ |
| | night rate vs. nitrate | /naitreit/ |
| | make lean vs. may clean | /meiklin/ |

Of importance here is that these pairs are clearly distinct in phonetic realizations even though they have identical phonemic sequences. The voiceless stops /p, t, k/ systematically realized as unreleased ones in word-final position, while they are strongly aspirated in word-initial position as can be seen in (2) (where C=means unreleased stop):

(2) [ʃip=reɪd] vs. [ʃip^hreɪd]
[naIt=reɪt=] vs. [naɪt^hreɪt=]
[meIk=lɪn] vs. [meɪk^hlɪn]

Therefore, Nakatani (1977, 1978) and Church (1987) show that the precise word boundary can be located by using these kinds of allophonic and prosodic cues; that is, in the above examples, word boundary will be located after unreleased stop and before the aspirated stop since unreleasing and aspiration are characteristic of word-(or syllable-)final and word-(or syllable-)initial of the stop consonants, respectively.

Furthermore, these kinds of allophonic cues and prosodic constraints are very helpful for parsing the utterance even in connected speech. For example, consider the following English sentence in (3) (cited from Church (1987), 180-182).

(3) a. Did you hit it to Tom?
b. [dɪdʒəhɪɹɪ?t^hət^ham]
c. [dɪdʒə # hɪɹ # ɪ? t^hə # t^ham]
(4) a. /h/ is "always" syllable initial,
b. [ɹ] is "always" syllable final,
c. [?] is "always" syllable final, and
d. [t^h] is "always" syllable initial.

As shown in Church, due to the English phonotactic and allophonic constraint in (4), the sentence (3a), with the narrow phonetic transcription (3b), can be parsed into the syllable size constituents as in (3c). After the utterance being parsed into manageable syllable size units, it will be much easier to match these units against the lexicon as illustrated in (5):

(5)

| parsed transcription | | decoding |
|---|---|---|
| [didʒə] | ::::::::::::::> | did you |
| [hiɹ] | ::::::::::::::> | hit |
| [ɪ?] | ::::::::::::::> | it |
| [t^hə] | ::::::::::::::> | to |
| [t^ham] | ::::::::::::::> | Tom |

The allophonic cue of voicing in Korean is also useful for parsing the utterance into phonological phrases. For example, consider the following phrases.

(6) a. 아버지가 방에 들어가신다.
(The father goes into the room.)
apʌci+ka # paŋ+e # tɨlʌkasinta.
father SUBJ room to goes into
[abədʒiga # paŋe # tɨrəgaʃinda]

b. 아버지 가방에 들어가신다.
(Father goes into the briefcase.)
apʌci (ka) # kapaŋ+e # tɨlʌkasinta.
Father (SUBJ) briefcase to goes into
[abədʒi # kabaŋe # tɨrəgaʃinda]

(7) a. 아기도 둑을 본다.
(The baby also sees the bank.)
aki+to # tuk+ɨl # ponta.
baby also bank OBJ sees
[agido # tugɨl # ponda]

b. 아기 도둑을 본다.
(The baby sees the thief.)
aki # totuk+ɨl # ponta.
baby thief OBJ sees
[agi # todugɨl # ponda]

The voiced stops, or the allophonic counterparts of voiceless unaspirated stops occur between the sonorant segments just within the same phonological phrases (cf. Park (1990)). Therefore, with the use of the allophonic cue of voicing, without relying on any higher level constraints, we can disambiguate the sentences in (6) and (7). That is, only voiceless stops can appear at the beginning of the phonological phrase. Hence, if voiced stops show up, then it cannot be at the beginning of the phonological phrase, but rather it must be within the same phonological phrase with the surrounding sonorant segments.

In sum, in the past it was believed that allophonic variation was not very useful, being considered as a source of noise, not a source of constraint. Hence, an aspirated /t/, a flap /t/, an unreleased /t/, etc., might be mapped into a single broad class of /t/, without any taking advantage of allophonic constraints. On the contrary, it has recently been shown that allophonic cues also provide important information about the suprasegmental context, which is very useful for parsing the utterances into prosodic units, along with other phonological cues such as phonotactic constraints, the sonority hierarchy, the maximal onset principle of syllabification, etc.

## Ⅲ. Revisiting Feature Approaches

One of the strong phonological arguments for the distinctive analysis over the segmental analysis is that the former is much better in describing natural classes and natural phonological rules, and thus capturing the significant phonological generalizations in languages. But, in most previous speech recognition models, segmental approaches had been preferred to the featural approaches, mostly due to the weakness of the distinctive feature theory *per se*. Hence, an aspirated /t/ or flap /t/ would be modeled with an atomic label like ASPIRATED-T or FLAPPED-T. However, as Church points out, this kind of modelling will require a large number of labels in order to represent all the allophones. For example, for an English phoneme /t/, at least the following atomic labels are needed such as ASPIRATED-T (*Tom, attack*), UNRELEASED-T (*pot, cat*), FLAPPED-T (*water, party*), RETROFLEXED-T (*treat, try*), ROUNDED-T (*twelve, twin*), PALATALIZED-T (*meet you*), UNASPIRATED-T (*stay, steam*), NASALRELEASED-T(*sweeten*), GLOTTALIZED-T (*button*), etc. Because as the number of labels increases, it becomes harder to work with all of these labels, Church prefers to view segments as a bundle of features. Furthermore, Many allophonic and phonological processes share the same environment: for instance, aspiration and retroflexing occur in syllable-initial position, while glottalization and unreleasing occur in syllable-final position. Hence, by utilizing these kinds of linguistic generalizations, we can make best use of allophonic cues for parsing the constituents in a bottom-up way without relying on higher level constraints.

In addition to these advantages, Church (1987) further shows that the feature analysis has advantages even in confusion error models. In most previous confusion models, either probabilistic or categorial ones, segments are treated as atomic objects. Of interest is, however, that most segmentation devices are more likely to make single feature errors (e.g. labeling a /t/ as a /d/ or /p/) than double feature confusion errors (e.g. labeling a /t/ as a /b/). For example, as can be seen in the following confusion matrix (8) (Seneff (1979), cited in Church (1987: 168)), single feature errors (36%, 62/173) are more probable than double feature confusion errors (5%, 9/173).

(8)

| | p | b | t | d | k | g |
|---|---|---|---|---|---|---|
| p | 19 | 4 | 4 | 2 | 4 | 1 |
| b | 3 | 9 | 0 | 1 | 0 | 3 |
| t | 4 | 0 | 39 | 1 | 4 | 0 |
| d | 4 | 3 | 9 | 10 | 1 | 0 |
| k | 4 | 1 | 2 | 0 | 17 | 8 |
| g | 1 | 1 | 0 | 1 | 5 | 8 |

Hence, if we compare this segmental confusion matrix (8) with the following feature level confusion matrix (9) (Church (1987: 171)), then we can find that, concerning stops, we may reduce the number of degrees of freedom at distinctive feature level rather than at the segmental level.

(9)

| a. | voiced | voiceless |
|---|---|---|
| voiced | 43 | 33 |
| voiceless | 22 | 76 |

| b. | labial | coronal | velar |
|---|---|---|---|
| labial | 35 | 7 | 8 |
| coronal | 11 | 59 | 5 |
| velar | 7 | 3 | 38 |

The phonemic confusion model in (8) has 36 degrees of freedom, while the feature based model in (9) has only 13 (4 voicing and 9 for place of articulation). By reducing the number of degrees of freedom, another benefit can be achieved; i.e., the parameters are easier to estimate because the rare events become less. Furthermore, in the feature analysis the recognition will be higher than in segmental analysis. For example, if we compare (8) with (9), as Church notes, we can also find that Seneff (1979) correctly recognized only 58% of the stops, but she made 68% and 73% correct recognitions for vocing and place respectively.

Finally, if the feature analyses are incorporated into a lexical organization, then it will be easier to recover from single feature errors. Suppose, for example, that the word *strength* are indexed with following three keys, i.e., as a sequence of manner, place, and voicing features.

(10) a. [fric] [stop] [son] [V] [nas] [fric] : manner
b. [cor] [cor] [cor] [V] [cor] [dent] : place
c. [vl] [vl] [vd] [vd] [vd] [vl] : voicing

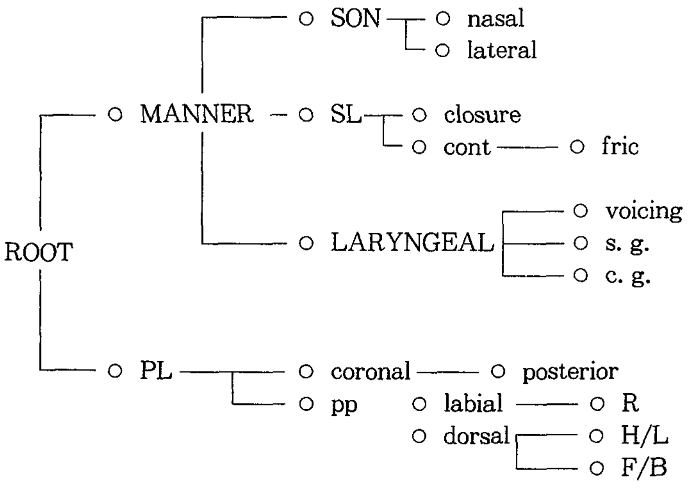
When the lexicon is accessed on all of these keys, words matching all three keys are more plausible candidates than those matching two or less keys. Thus, if we adopt a probabilistic framework, then we may ignore words with two or more feature errors since there is little hope of recovering. Furthermore, of interest is that manner acoustic cues, are in general very easy to tell in spectrogram. Hence the manner detector usually gets higher rate of correct recognition than the place detector. So if we use manner keys as the first screening for matching, we can not only save time but also increase matching efficiencies.

From the above Church's arguments and others, we may draw a conclusion that the feature analysis is preferred to the segmental analysis in speech recognition. If then, it naturally follows to ask a question why most speech models have adopted the segment based approaches instead of the feature based ones. It is probably mainly due to the weaknesses of the distinctive feature theory in early generative phonology *per se*. First of all, all the features do not have direct acoustic correspondences. For example, labial and velar consonants are specified as [-cor, +ant] and [-cor, -ant] respectively for their place features. But what are the acoustic cues for [-cor], [+ant], and [-ant]? From the acoustic point of view, as for the place distinction, the distinctive features of 'labial', 'coronal', and 'dorsal' using Prague School's notion of privativeness is better than the binary use of 'anterior' and 'coronal'. (As for other phonological weaknesses of traditional feature matrix system such as the inabilities of representing contour segments like the affricate /tʃ/ or prenasalized stop /nt/ and complex segments like /kp/, see Sagey (1986) and Kim (1987, 1990).) In what follows, I argue that the feature theory should be revised with the theory of feature hierarchy, using the notion of feature privativeness.

## Ⅳ. Feature Hierarchy

In the framework of nonlinear phonology, Clements (1985, 1989), Sagey (1986), Kim (1987, 1989, 1990), McCarthy (1988), Avery & Rice (1989), and among others recently proposed that distinctive features be organized hierarchically, since certain groups of features consistently behave as a unit with regard to certain phonological processes such as assimilation and resequencing. For instance, many languages show place assimilation by which place features [$\alpha$ coronal, $\beta$ anterior, $\gamma$ high, $\delta$ back] behave as a unit, and thus features relevant to the place of articulation should be grouped under the node of PLACE. Similarly, features relating to the state of glottis such as [spread glottis], [constricted glottis], and [voicing] should be organized under the LARYNGEAL manner node. I propose the following (tentative) modified version of feature hierarchy, based on both articulatory and acoustic grounds. (Compare with Clements (1985, 1989), Sagey (1986), Kim (1987, 1990), McCarthy (1988), among others.)

(11)




(SON=SONORANT, SL=SUPRALARYNGEAL,
s. g.=spread glottis, PP=PERIPHERAL,
c. g.=constricted glottis, R=round,
H=high, L=low, F=front, B=back)

In this feature hierarchy, the root node (representing the existence of the segment) is divided into two class nodes MANNER and PLACE. The MANNER node is further divided into SONORANT node dominating nasal and lateral features, and SUPRALARYNGEAL MANNER node dominating the degree of closure features, and LARYNGEAL MANNER node dominating glottal status features. The PLACE node dominates place feature coronal and PERIPHERAL place features of labial and dorsal, which in turn dominate their own relevant features.

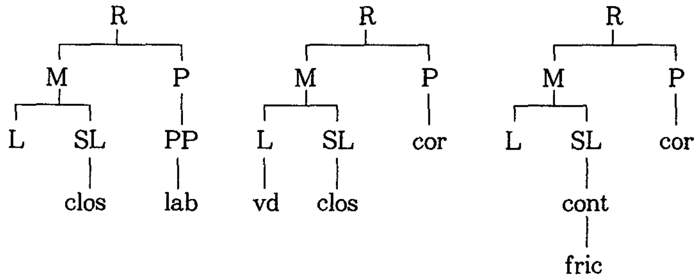
One of the most significant characteristics in this feature hierarchy is that each feature has corresponding articulators and acoustic cues. For example, the feature nasal has an articulatory command of lowering the soft palate in order for the air to pass through the nasal cavity, and corresponding acoustic nasal formant structures (usually at 250, 2,500, and 3,250 Hz). Similarly, lateral sounds are produced by lowering the mid section of the tongue at both sides or at only one side, thereby allowing the air to flow out of the mouth in the vicinity of the molar teeth, and they also have corresponding lateral acoustic formant structure (similar to that of vowels but with formants in the neighborhood of 250, 1,200, 2,400Hz and the higher formants with highly reduced intensity). As for the LARYNGEAL manner features, voicing has a command of vibrating the vocal cords and the corresponding acoustic cue (little bars in very low frequencies). Similarly, the feature spread glottis not only has an articulatory command of spreading glottis, but it also has a corresponding acoustic cue of noises in higher frequencies. As for the PLACE features, each feature has its own articulators. That is, as the active articulators, coronal involves tongue tip or tongue blade, while labial and dorsal sounds involve the lip and tongue body respectively. These place features have also their own acoustic cues and thus we can tell the differences among them with their cut-off frequencies and the onset/offset formant structures ($F_1$, $F_2$, $F_3$) of the following/preceding vowels.

Another characteristic in this feature hierarchy is that the dependency relations among features are clearly expressed. For example, as can be seen in (11), the feature 'labial', capturing the fact that rounding is relevant only to the lips, but neither to the tongue tip nor to the tongue body. Similarly, the feature 'nasal' and 'lateral' are dependent on SONORANT feature node, since only nasal and lateral sounds among consonants will have sonorant formant structures. On the other hand, the oral stop feature 'closure' and the fricative feature 'frication' are dependent just on SUPRALARYNGEAL manner node, but not on SONORANT node, and thus they lack the sonorant formant structures. The feature 'posterior', which is for the palatal sounds, is dependent on coronal node, since palatal sounds also involve the tongue blade as their active articulator like alveolar sounds.

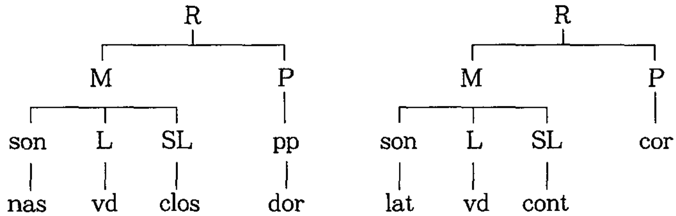
Another important characteristic is that each features are used with privative opposition: that is, the specification of features represents the presence of the relevant features, while nonspecification of the feature represents the lack of relevant articulatory and acoustic cues for the segment in question. Hence, for example, nasal sounds will have both the SONORANT node and its dependent feature 'nasal' specified, but oral sounds will not be specified for the feature 'nasal' since oral sounds do not involve soft palate lowering nor have corresponding acoustic nasal formants. As for the node SONORANT, sonorant segments such as vowels, glides, nasals, and liquids will be specified as sonorant, since all these segments have formant structures in common. On the other hand, for obstruents such as stops, fricatives and affricates, the feature sonorant will not be specified since obstruents do not have any formant structures at all. As for the place features, coronal consonants have only coronal feature specified without any specification of labial and dorsal, since only the tongue tip or blade is involved, and neither the lips nor the tongue body is relevant to producing coronal sounds. Similarly, labial consonants will have only the labial specified, while velar consonants only the dorsal specified. Thus the feature specification of any distinctive feature (or class feature node) implies that the segment in question involves the corresponding feature (either articulatorily or acoustically, or both), while nonspecification indicates the irrelevance of the corresponding feature.

Hence, examples such as voiceless labial stop /p/, voiced coronal stop /d/, voiceless coronal fricative /s/, the velar nasal /ŋ/, and lateral /l/ will be represented as follows:

(12) a. /p/ b. /d/ c. /s/



d. /ŋ/ e. /l/




As can be seen in (12), obstruents are not specified for the feature SONORANT and its dependent features of 'nasal' and 'lateral' (as in 12a-c), since they lack the formant structures. With regard to the LARYNGEAL manner features, the feature 'voicing' will be specified only for the voiced segments, while the feature 'spread glottis' specified for aspirated segments. But nothing will be specified for the voiceless unaspirated obstruents, since they involve neither voicing nor aspiration articulatorily as well as acoustically (as in 12a, c). Concerning the place of articulation, coronal segments are specified only with 'coronal', without any specification of 'labial' and 'dorsal' (as in 12b, c, e). But the labial sound /p/ is specified with PERIPHERAL and 'labial' (in 12a), while velar nasal /ŋ/ is specified with PERIPHERAL and 'dorsal' (in 12d). In short, only relevant features and feature nodes for the segment in question will be specified, while irrelevant features will not be specified.

## V. Privativeness and Phonetic Underspecification

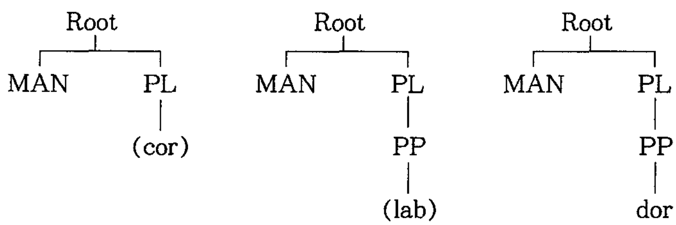
In early generative phonology, it has been agreed that phonological representations may be underspecified but surface phonetic representations should be fully specified by either redundancy rules or phonological rules (Archangeli (1984), (Chomsky & Halle (1968)). This requirment of the fully specified phonetic representation is probably another reason why the feature approaches have been neglected in speech recognition.

As shown in previous section, the notion of privativeness is very important in the theory of feature hierarchy: the specification and nonspecification of each feature represent the presence and absence of the relevant cues in question. Hence, the specification of the feature node SONOR-

ANT represents the existence of formant structures, typical for sonorant segments such as vowels, glides, nasals, and liquids. In contrast, the nonspecification of SONORANT represents the absence of the formant structures, typical for obstruents such as stops, fricatives and affricates. What we should note is that this privativeness notion in feature hierarchy will naturally lead to surface underspecification.

Before further pursuing phonetic underspecification, let us first consider phonological underspecification. In general, most arguements for underspecification come from phonological phenomena of transparency and unidirectional behaviors in the assimilation processes. Due to the page length limit, I will consider just the place features. On the basis of language inventories, language acquisition, and language changes, the coronal segment is regarded as the least marked one among coronal, labial and velar sounds (Kean (1975), Kim (1987), Avery & Rice (1989)). Under the theory of underspecification, the unmarked feature will not be specified underlyingly, but will be supplied later by redundancy rules. Thus coronals will not be specified for their place feature 'coronal', as in (13a) below, where parenthesis indicates underspecification. Between labial and velar segments, both specified as PERIPHERAL, labials are regarded as less marked than velars and thus will not be specified for their peripheral feature as in (13b), while velars will have peripheral feature 'dorsal' specified as in (13c):

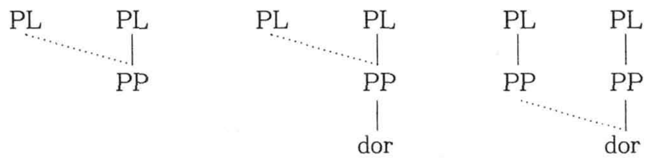
(13) a. /t/ b. /p/ c. /k/




In certain languages, coronal sounds show an asymmetrical behavior of transparency with regard to the vowel harmony processes. Unlike labial or velar consonants, vowels separated by coronals often behave as if they are adjacent. In other words, coronal sounds are transparent concerning the spreading of the adjacent vowel feature spreading. This transcoronal vowel spreading can be accounted for by the underspecification of coronal node in underlying phonological representation.

With respect to the place assimilation, coronal sounds usually behave asymmetrically. For example, in Korean, coronals optionally assimilate to the place of the following labials and dorsals (e.g. /kunpam/ → [kumbam] 'roast chestnut', /kunkokuma/ → [kuŋgoguma] 'roast sweet potato'), but not vice versa (e.g. /kaŋto/ → *[kando] 'a robber', /pamnun/ → *[pan nun] 'night vision'). Between the peripheral sounds, labials optionally assimilate to the following velars (e.g. /kamki/ → [kaŋgi] 'cold'), but not vice versa (e.g. /kaŋmul/ → *[kammul] 'the river'). This asymmetrical unidirectional behavior of place assimilation in Korean can be properly accounted for as spreading the specified, or marked place feature to the unmarked one, as illustrated in (14).

(14) a. alv > lab b. alv > vel c. lab > vel




By allowing the unmarked segments to be underspecified underlyingly, the peculiar behaviors such as phonological transparency and unidirectional assimilation can be easily accounted for. Now then how can the theory of underspecification be implemented into speech recognition? As for place assimilation, the direction of assimilation is unidirectional, i.e., the unmarked one to the marked one. Hence, for example, if the unmarked coronal sequence such as [tando] was found, the matcher will try to find only the unmarked sequence like /tanto/ 'a dagger', but not a sequence of marked and unmarked ones such as /tamto/ or /taŋto/. On the other hand, if the sharing of the marked features such as [kumbam] was found in the spectrogram, the matcher will interpret it as either a sequence of identical marked ones like /kumpam/ or a sequence of the unmarked and the marked ones like /kunpam/ by spreading of assimilation.

Now let us go back to the phonetic underspecification. If certain features are regarded as the unmarked ones, then these features will not be specified underlyingly. But in surface phonetic representation, these unmarked features, after being supplied by redundancy rules, will show up. For example, coronal segments are underspecified for their place feature 'coronal' as in (13a), and thus they are subject to assimilation towards the marked segments as illustrated in (14). However, if there is no assimilation, then the redundancy rule will supply the feature 'coronal' as the default place feature for the unspecified consonants. But, as I argued in my thesis (Kim (1987)), the difference between underspecification and nonspecification should be noted here. If certain features are underspecified, then redundancy rules will supply the default features to these underspecified segments. In contrast, if they are nonspecified, then the segment in question will remain phonetically as unspecified unless the features are supplied by the phonological processes such as assimilation. For example, unlike the feature coronal, the peripheral features labial and dorsal are nonspecified for the coronal segments. Therefore, these features will remain as unspecified unless they are supplied by spreading. Since they are used as privative ones, these features will never be involved articulatorily in producing the coronal sounds, and thus naturally there will be no corresponding acoustic cues.

Furthermore, as argued before, the class node can also be nonspecified. For example, the segment /h/, usually regarded as the voiceless counterpart of the vowel, will not have the node PLACE specified. The peculiar behaviors of /h/ in Korean, then, will be accounted for by this lack of the node PLACE (for further details, see Kim (1987, 1990)). In word-initial position, the segment /h/ usually spells out as having the same place of articulation of the following vowels, which can be accounted for by spreading of the place node of the following vowel to the unspecified place node. Recently, Keating (1988) further argues that in connected speech the intervocalic /h/ in English does not get the place features by spreading, but it simply interpolate between the surrounding vowels. Her argument is as follows: If it is due to spreading then it will share the same phonetic properties with the following vowel for most of its duration. But the acoustic evidence of the intervocalic /h/ shows that the assimilatory effect on the /h/ is a dynamic, transitional one, not a static one. From this lack of steady period of sharing, she draws a conclusion that /h/ starts out unspecified for place features, and it remains unspecified. In sum, due to the privative use in the feature hierarchy, certain features or feature nodes may not be specified underlyingly and they may remain as unspecified phonetically.

## Ⅵ. Conclusion

In this paper, I show not only the usefulness of variant cues for parsing without relying on the higher level constraints but also the superiority of the featural anaylsis to the segmental one. Since the reason why the feature analysis has been neglected in speech recognition is mainly due to the defects of the feature theory *per se*, I argue that the binary notion of feature values and the requirement of fully specified phonetic representation should be given up, and instead the privativeness notion of features be adopted, which leads to surface phonetic underspecification. This revised feature approach is superior to the segment approach in speech recognition since it not only makes use of both invariant and allophonic cues, still utilizing the signficant linguistic generalizations, but it also provides a more robust estimation of confusion probabilities.

## References

Archangeli, D. (1984) 'Underspecification in Yawelmani Phonology and Morphology,' Ph. D. Diss., MIT.

Arery, P. and K. Rice(1989) 'Segment Structure and Coronal Underspecification,' *Phonology* 6, pp. 179-200.

Campbell, L. (1974) 'Phonological Features: Problems and Proposals,' *Language* 50, pp. 52-65.

Chomsky, N. and M. Halle (1968) *The Sound Pattern of English*, Harper & Row.

Clements, G. (1985) 'The Geometry of Phonological Features,' *Phonology* 2, pp. 225-252.

Clements, G. (1989) 'A Unified Set of Features for Consonants and Vowels,' ms. Cornell University.

Cole, R. and J. Jakimik (1980) 'A Model of speech Perception,' in R. Cole (ed.), *Perception and Production of Fluent Speech*, Lawrence Erlbaum, Hillsdale, N. J.

Church, K. (1987) *Phonological Parsing in Speech Recognition*, Kluwer Academic Pub., Boston.

Kean, M. (1986) 'The Theory of Markedness in Generative Grammar,' Ph. D. Diss, MIT.

Keating, P. (1988) 'Underspecification in Phonetics,' *Phonology* 5, pp. 275-292.

Kim, Kee-Ho (1987) 'The Phonological Representation of Distinctive Features: Korean Consonantal Phonology,' Ph. D. Diss., University of Iowa.

Kim, Kee-Ho (1989) 'Some Phonological Implications of Underspecification in Feature Hierarchy,' Linguistic Research (Kyung Hee Univ.) 10, pp. 145-161.

Kim, Kee-Ho (1990) 'Nonspecification and Underspecification in Feature Hierarchy,' *Linguistic Journal of Korea* 15, pp. 153-193.

Klatt, D. (1980a) 'SCRIBER and LAFS: Two New Approaches to Speech Analysis,' in W. Lea (ed.), *Trends in Speech Recognition*, Prentice-Hall.

Klatt, D. (1980b) 'Overview of the ARPA Speech Understanding Project (1),' in W. Lea (ed.), *Trends in Speech Recognition*, Prentice-Hall.

Ladefoged, P. (1982) *A Course in Phonetices*, Harbourt, Brace, N. Y.

McCarthy, J. (1988) 'Feature Geometry and Dependency: A Review,' *Phonetica* 43, pp. 84-108.

Nakatani, L. and K. Dukes (1977) 'Locus of Segmental Cues for Word Juncture,' *JASA* 62, pp. 714-719.

Nakatani, L. and J. Schaffer (1978) 'Hearing 'Word' without Words: Prosodic Cues for Word Perception,' *JASA* 63, pp. 234-245.

Paradis, C. and J. Prunet (1989) 'On Coronal Transparency,' *Phonology* 6, pp. 317-348.

Park, D-S. (1990) 'Lexicon and Syntax in Korean Phonology,' Ph. D. Diss., Univ. of Hawaii.

Sagey, E. (1986) 'The Representation of Features and Relations in Nonlinear Phonology,' Ph. D. Diss., MIT.

Seneff, S. (1979) 'A Spectrogram Reading Experiment,' Unpublished Paper, MIT.

Shoup, J. (1980) 'Phonological Aspects of Speech Recognition,' in W. Lea (ed.), *Trends in Speech Recognition*, Prentice-Hall.

Stevens, K. (1981) 'Invariant Acoustic Correlates of Phonetic Features,' *JASA* 69, suppl. S31.

Steriade, D. (1987) 'Redundant Values,' *CLS*, Part 2, pp. 339-362.

Zue, V. (1983) *Speech Spectrogram Reading: An Acoustic Study of English Words and Sentences*, Andover, Ma.

Zue, V. and R. Schwart (1980) 'Acoustic Processing and Phonoetic Analysis,' in W. Lea (ed.), *Trends in Speech Recognition*, Prentice-Hall.

Department of English
Korea University
Anam-dong, Sungpuk-ku
Seoul 136-701
Korea