An Optimality Theoretic Approach to the Feature Model for Speech Understanding*

Kee-Ho Kim**

<Received: September 15, 1997>

ABSTRACT

This paper shows how a distinctive feature model can effectively be implemented into speech understanding within the framework of the Optimality Theory(OT); i.e., to show how distinctive features can optimally be extracted from given speech signals, and how segments can be chosen as the optimal ones among plausible candidates. This paper will also show how the sequence of segments can successfully be matched with optimal words in a lexicon.

Key words: distinctive features, speech understanding, Optimality Theory

1. Introduction

After the introduction of the spectrogram in the late 1940s (Koenig *et al.* 1946), many research studies have been done on the acoustic cues for distinctive features. However, the feature-based knowledge model in general does not succeed in drawing strong attractions in speech understanding, due to its lack of success in the practical speech recognition system. Rather, the statistical pattern-recognition approach, in which no speech-specific knowledge is explicitly used, has been favored due to its simplicity of use and relatively high performance, etc. (Rabiner & Juang 1993).

This paper assumes the hybrid type of artificial neural networks, which incorporates both acoustic phonetic knowledges and other higher level linguistic knowledges, also using statistical methods whenever necessary. This paper will show how the feature-based acoustic knowledge model can be effectively implemented in speech understanding. This

---

* This is the slightly revised version of the paper presented in ICSP'97 (1997. 8. 27), and has been supported by the faculty overseas research grant from Korea Research Foundation.

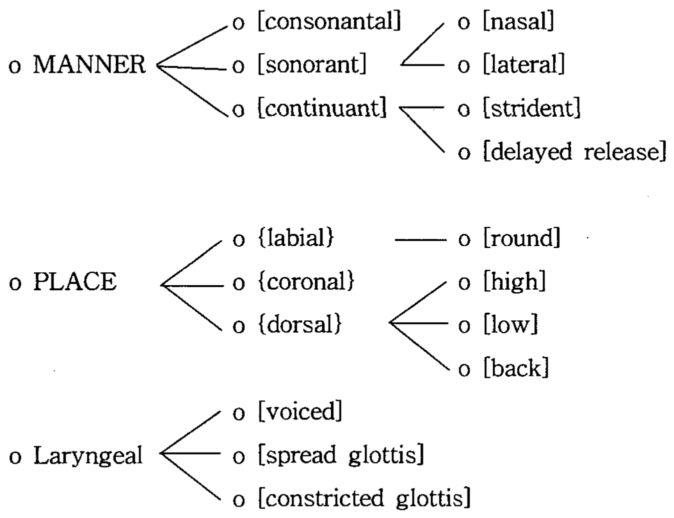
** Department of English, College of Liberal Arts, Korea University.

paper is organized as follows. In section 2, after briefly introducing a hierarchical model of feature geometry, the key acoustic characteristics of distinctive features will be presented. In section 3, some of Liu's (1995, 1996) experiment results showing that certain acoustic cues can successfully be detected by the automatic landmark detection algorithm will be recapitulated. And then, based on the multi acoustic correlates of features, this paper will show how these distinctive features can optimally be extracted from given speech signals in the OT framework. Once the relevant features are extracted from the given input, the optimal candidate will be selected by a screening of these features. Finally, in section 4, this paper will show how prosodic features can be manipulated for phonological parsing, leading to an efficient lexical retrieval.

2. Acoustic cues for distinctive features

Since the introduction of distinctive features in Jakobson *et al.* (1952/63), so far a variety of feature models have been proposed. In this paper, based on Sagey (1986), Kim (1991) & Halle & Steven (1991) etc., a hierarchical feature model for speech recognition will be presented as in (1), in which features are divided into the major feature groups Manner, Place, & Laryngeal.

(1) hierarchical feature model



These feature segments are divided into major classes of vowels, sonorant consonants, fricatives and stops, through the primary features of [consonantal], [sonorant], and [continuant], which generally have more salient acoustic cues than place or articulator-dependent features. What follows is a description of major classes of segments and the acoustic characteristics of distinctive features in the feature model (1). Here the acoustic cues in Jakobson *et al.* (1952/63), Stevens (1992, in print), Kent & Read (1992), Liu (1995), Kent, Dembowsky & Lass (1996), etc. will be recapitulated.

Vowels generally constitute the nucleus of the syllable and they are distinguished from consonants with the feature [consonantal].

- [-consonantal] : the maximum amplitude in F_1, overall steady high spectrum energy.
- [+consonantal] : abrupt discontinuity in spectrum. In complete closure, at least 15-20 dB abrupt spectral energy change during 10 ms interval.

Vowels are divided by the tongue dorsal features [high], [low], [back], and the lip shape feature [round].

- [+high] : lower F_1, small difference between F_1 and F_0 (less than 3 Bark). (High vowels, on the average, have a higher F_0, lower intensity, shorter duration, lower F, and lower F - F_0 difference than low vowels.)
- [+low] : higher F_1, larger difference between F_1 and F_0 (more than 3 Bark).
- [-high, -low] : a rather uniform spacing of formants.
- [-back] : relatively large separation between F_1 and F_2 and a relative closeness of F_2 and F_3. (Front vowels have larger $F_2 - F_1$ but smaller $F_3 - F_2$ difference than back vowels.)
- [+back] : small separation between F_1 and F_2, & large separation between F_2 and F_3.
- [+round] : lowering of all formants. (Round vowels have larger sum of (F + F_2 + F_3) than unround vowels.)

Glides are [-consonantal], but they do not constitute the nucleus of the syllable. Glides show gradual formant transitions1) (about 75 to 150 ms, compared with stops (50-75 ms)), and their amplitude of low frequencies is relatively low to the adjacent vowels. The F_2 and

---

1) The experiment of speech recognition(Liberman *et al.* 1956) shows that if the transition period is short (less than 40 ms), the segments in question are perceived as stops while the transition is longer than 40 ms, they perceived as glides.

F_3 loci for /w/ are about 800 Hz and 2,200 Hz respectively, and its landmark is represented by the lowest F_2 frequency point. The F_2 and F_3 loci for /j/ are about 2,200 Hz and 3,000 Hz respectively, and its landmark is represented by the highest F_3 peak point.

Sonorant consonants are distinguished from obstruents by the feature [sonorant].

- [+sonorant] : the existence of formants in spectrogram.

- [+lateral] : low F_1 frequency, abrupt spectral energy increase in F_3 - F_4 frequencies at the explosion.

The liquid /l/ has a falling F_2 and steady F_3, while the English /r/ shows abrupt falling transitions of both F_2 and F_3. Nasals are further distinguished by nasal murmur acoustic cues.

- [+nasal]: nasal murmur (low-frequency energy, often less than 0.5 kHz (for male), voicing bar, place formant transition to vowels similar to stops.

Nasal vowels or nasalized vowels also show the following acoustic cues: lower frequency nasal formant, reduction of overall energy, increase of the formant bandwidths. In general, [+nasal] vowels have larger formant bandwidths and higher F but less intensity and less sum of (F_2 + F_3) than [-nasal] vowels.

Obstruents are divided into stops and fricatives by the feature [continuant].

- [-continuant] : silence (for voiceless stops) or a low energy interval, or the voice bar (for voiced stops), 5 to 15 ms brief acoustic burst (for syllable-initial or -medial stops) and 20 to 40 ms transient. At least 10-15 dB abrupt energy increase or decrease in the mid- and high-frequencies at implosion and explosion. The sonorant stops (or nasals) also show an energy decrease, e.g. /m/ shows about 17 dB F_2 decrease for the peak (20 ms interval) and 10 dB F_1 decrease.

In general, stops have a silence - burst - aspiration (for voiceless one) - transition sequence in syllable-initial position, but a transition - silence - (optional) burst - (optional) aspiration sequence in syllable-final position. They can further be distinguished by the following place-of-articulation acoustic cues:

- {coronal} : i) coronal burst spectrum, or energy prominence around 4 kHz (diffuse, flat or rising spectrum template), ii) formant transition (F_1 increases from a near-zero value to the frequency for the vowel; F_2 begins from a value close to the assumed F_2 locus of 1,800 Hz, and the direction of F_2 shift depends on the F_2 of the adjacent vowel.

- {labial} : i) labial burst spectrum (diffuse, flat or falling spectrum template). ii) formant transition (same as for coronal F_1, F_2 increases from a low frequency of about 800 Hz to the F_2 of the vowel. F_3 locus is low compared with that of the vowel.)
- {dorsal} : i) velar burst spectrum (compact (mid-frequency emphasis) spectrum template), ii) formant transition (same as for coronal F_1, F_2 has at least two loci, one at a low frequency of about 1,300 Hz and the other near 2,300 Hz. The F_2 and F_3 formant frequencies are initially very close but diverge during the transition.)

Fricatives have a salient acoustic cue of aperiodic noise, and its place can be distinguished by the following acoustic cues:

- [+continuant] : (fricatives) : frication (aperiodic noise) and transition.
- [+strident] : a very intense noise energy.
- {labial} : /v, f/ low-energy (weak), diffuse (flat) spectrum; F_1 increases, and F_2 increases except for some back vowels.
- {coronal} : /s/ intense noise with most energy in the high frequency (above 4 kHz), high mean and a marked skewness for spectrum; /ʃ/ intense noise with most energy in mid to high frequencies (about 2 kHz).
- {glottal} /h/ : low energy (weak) diffuse (flat) noise spectrum, little formant change.

Affricates have the acoustic cues of a stop followed by a fricative. However, affricates also have a short rise time, or a relatively rapid increase in noise energy. On the other hand, the onset of noise in fricatives is gradual, and they a longer noise interval than affricates.

- [+delayed release] (for affricates) : stop gap (silence or low-energy interval) and frication.

The acoustic cues for the laryngeal features are as follows:

- [-voiced] : (for stops) In initial position, long VOT, F_1 cutback or a delay in F_1 relative to the higher formants, the higher F_0 of the following vowel than in [+voiced] stop; in medial position, no voicing bar, higher F_0 of the following vowel and longer closing duration than voiced one.

- [+aspirated] : longer VOT, aperiodic noise.
- [+constricted]: shorter VOT, longer closing duration.

On the basis of these acoustic characteristics of distinctive features, in what follows, how these acoustic cues can be manipulated to search for the optimal features in the OT framework will be shown.

3. Feature Detection in Optimality Theory

When speech signals are given as an input, one of the major initial steps of speech recognition in the knowledge-based model is to detect its featural cues. Liu (1995, 1996) recently presented an algorithm for automatically detecting acoustically abrupt landmarks such as *abrupt-consonantal*, *abrupt*, & *nonabrupt*. Liu also showed how landmark detection can be used to search for distinctive features. According to Liu's algorithm, certain acoustic characteristics will be measured to locate a sequence of landmarks, which will lead to a specification of features: For example, band 1 (0.0 - 0.4 kHz) energy and band 1 ROR (rate-of-rise) properly provides the landmark for *voiceless*. The *sonorant* landmark (reduced energy in bands 2-5 (1.3 - 8 kHz)) properly detects the [+consonatal, +sonorant]. The silence interval followed by a sharp increase in energy in high frequencies acoustically correlates with (+)*burst* landmark, which signifies the [+aspirated] stop or [+delayed release] affricate burst, while the sharp decrease in high frequency energy followed by a silence interval correlates with (-)*burst* landmark, which signifies the offset of frication or aspiration noise due to a stop closure, and so on.

Of interest in this study is that Liu's landmark detection experiments on clean speech showed relatively low error rate results: i.e., it was reported that the *glottal* landmarks, corresponding to stops and fricative among others, had a detection rate of 98%, while the *burst* landmarks2), corresponding to bursts and the cession of noise, had a 90% detection rate. Particularly, the voiceless obstruents were reported to have a near 100% *glottal* detection rate. Liu's landmark detector generally performed with a detection rate of about 90%. In short, Liu's landmark detection experiments proved that certain distinctive features such as [consonantal], [continuantinuant], [sonorant], [voiced], etc. could easily be extracted through the successful process of locating the landmarks.

However, Liu's landmarks do not cover all the distinctive features in our feature model (1). In order to detect all the relevant features, the ideal speech recognition algorithm should

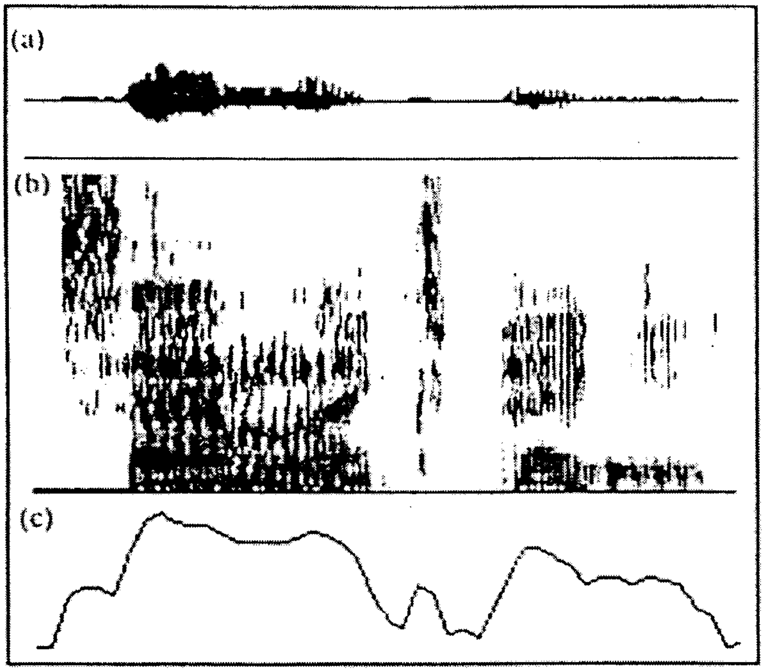
---

2) Note here that since burst landmarks can occur during voiceless period, only the regions delimited by a (-) glottal on the left and a (+) glottal on the right are searched.

be accommodated to be able to manipulate other acoustic measures such as zero-crossing rate, formant transition, peak FFT amplitude, overall energy spectrum, spectrum amplitude change rate, etc., along with Liu's detecting means such as spectrogram and 6 frequency bands analyses (for landmark detection) (cf. Johnson 1994).

Furthermore, as noticed, phonetic features are associated with one or more acoustic cues. This multi-cue relationship plays an important role in optimal feature searching in the framework of OT: the more acoustic cues detected, the more optimal candidates selected. In order to see this, let the following Figure 1 be considered, which shows the waveform, spectrogram and overall spectral energy of 'Sam likes Peggie.'

Figure 1. The waveform (a), spectrogram (b) and overall spectral energy (c) of 'Sam likes Peggie.'



As can be seen in Figure 1, each segment has its own distinctive acoustic cues: the segment /s/ shows an intense fricative noise; segments /æ, aɪ, ɛ/, high spectrum energy and vowel formants; /p/, voiceless, aspirated, labial stop burst and formant transition; and so on. (For further details, refer the acoustic cues in section 2.) To make it clear how multi cue relation works in OT, take a close look at the segment /s/ in the spectrogram. As we can see, there exists several acoustic cues for the feature [+consonantal]; i.e., a very high z.c.r., no voice bar, low spectrum energy, etc., and above all, the salient high frequency noise energy. In the framework of OT (Prince & Smolensky 1993, McCarthy & Prince 1995, Hayes 1996), the selection of the feature [+consonantal] over [-consonantal] can be accounted for as in Tableau 1.

Tableau 1.

	I-noise	Noise	high-zcr	voiceless	low-spE	abSEch
[-consonantal	*!	!	!	!	!	!
@ [+consonantal]	@	@	@	@	@	@

(*I-noise*: intense high frequency noise energy; *Noise*: aperiodic noise energy; *high-zcr*: high zero crossing rate; *voiceless*: no glottal vibration; *low-spE*: low spectrum energy; *abSEch*: abrupt spectrum energy change; *: negative; *!: critically negative; @: positive.)

Note here that each acoustic cue plays a role as an acting constraint to detect distinctive features: i.e., whenever a certain phonetic cue is detected, it activates as a positive, negative, or neutral constraint, depending on the relevance of the acoustic cue. So, whenever the acoustic cue of intense noise energy is detected, the *I-noise* constraint will activate to give a positive value for [+cons] but a negative value for [-consonantal]. In the same way, whenever noise energy, high z.c.r., voiceless, low spectrum energy, etc. are detected, these cues will automatically assign the proper values for [consonantal]. So [+consonantal], which has six positive values, will be optimally selected over [-consonantal], which has six negative values. In OT, the more positive values they get, the more optimal they become.

Similarly the detection of the acoustic cues such as noise energy, high z.c.r., voiceless, etc., along with the nonexistence of formants (*no-For* constraint) will make [-sonorant] optimally selected over [+sonorant], as illustrated in tableau 2.

Tableau 2.

	I-noise	Noise	no-For	voiceless	low-spE	high-zcr
[+sonorantl	*!	!	!	!	!	!
@ [-sonornatl]	@	@	@	@	@	@

In the same vein, [+continuant] and [+strident] will be selected over [+continuant], [-strident] for the segment /s/, as illustrated in tableaux 3 and 4.

Tableau 3.

	I-noise	Noise	high-zcr
[-continuant]	*!	!	!
@ [+continuant]	@	@	@

Tableau 4.

	I-noise	Noise	high-zcr
[-strident]	*!	@	@
@ [+strident]	@	@	@

Of note here is that unlike the constraints in the previous phonological theories, the constraints in OT are violable and hierarchically rankable. This *violability* and *rankability* of constraints, which consists of the core of the OT theory, plays a very important role in speech recognition. As everybody acknowledges, one of the main difficulties in continuous speech recognition lies in segmentation and labeling, due to insertion, deletion, or featural overlapping, which stem from phonological assimilation, etc. These problems become much less problematic in OT, due to constraints' violability.

As already noticed, there exists at least one or more acoustic cues for each distinctive feature, and the optimal selection of the feature is based on these acoustic cues or constraints, and these constraints can be violated. Hence, missing errors of detecting certain acoustic cues are allowable in OT. In other words, even though certain acoustic cues are not detectable due to various reasons, the other remaining cues are sufficient enough to search for the optimal features. For example, even though the acoustic cue of the high zero crossing rate is missed, i.e. *high-zcr* constraint is violated, [+consonantal] and [-sonorant] will be optimally selected over [-consonantal] and [+sonorant] due to other acoustic cues of intense high frequency noise (*I-noise* constraint) and no glottal vibration (*voiceless* constraint), etc.

The constraints' rankability in OT is also efficient in speech recognition to save searching time, by putting the more salient cues higher than less salient, or redundant cues. For example, among the various acoustic cues of the voiceless alveolar fricative [s], the most salient cue would be the intense high frequency (above 4.0 kHz) noise energy. Therefore, once the salient strident cue is detected, we do not need to spend time searching for other acoustic cues such as low spectrum energy, no formants, etc. to find the optimal [+consonantal] and [-sonorant], since they are redundant, or predictable from the strident cue. In OT, this efficiency in saving time can easily be captured by putting the salient cue, or *I-noise* higher than other constraints, as illustrated in tableaux 1-4.

The optimal feature values for the vowel [æ] in 'Sam', e.g. [-consonantal, +sonorant] can be extracted by the acoustic cues of the maximum amplitude in F_1, overall high spectrum energy, and the existence of formants, etc., as illustrated in tableaux 5 and 6.

Tableau 5.

	Max F_1	HighSepecE	Formants
[-consonant]	*!	!	!
@ [+consonant]	@	@	

Tableau 6.

	Max F_1	HighSpecE	Formants
[-sonorant]	*!	!	!
@ [+sonorant]			@

As seen in tableaux 5 and 6, the detection of acoustic cues of maximum F_1 energy and overall high spectrum energy in the spectrogram will automatically activate *MaxF* and *HighSpecE* constraints to give a positive value for [-consonantal], but a negative value for both [+consonantal] and [-sonorant]. Also by detecting the existence of formants, the constraint *Formants* will activate to give a positive value for both [-consonantal] and [+sonorant] but a negative value for [-sonorant], while it will be neutral for [+consonantal]. Hence, by manipulating the acoustic cues as filters, the optimal features can be selected; i.e., the one that has more positive values wins as an optimal one over the one that has negative values.

Once the relevant features are extracted from the given input, the optimal candidate segment will be selected by screening the combination of features in the same vein. For example, when the input signal for [s] passes through the feature detection process, it will have [-consonantal, -sonorant, +continuant, +strident, -voiced, ...] as its optimal output features. In our OT framework, the manner features, which divide segments into major groups, will be ranked higher over place features, due to its relatively salient acoustic characteristics. The selection of /s/ as an optimal one over /z/, /f/, /t/, /n/, /l/, vowels, etc. will be accounted for as illustrated in tableau 7.

As seen in tableau 7, all vowels are rejected by violating the constraint [+consonantal], [-sonorant], [+strident], etc., and nasals and liquids are disqualified through the violation of the constraint [-sonorant], [+strident], and [-voiced], and stops are disqualified by the constraint [+continuant]. In the same way, fricatives /f/ and /z/ are rejected by the constraints of [+strident] and [-voiced] respectively. Hence, by scanning the constraints of [+consonantal], [-sonorant], [+continuant], [+strident], and [-voiced], /s/ will be selected as the optimal segment.

Tableau 7.

	[+consonatal]	[-sonorant]	[+continuant]	[+strident]	[-voiced]
@ /s/	@	@	@	@	@
/z/	@	@	@	@	*!
/f/	@	@	@	*!	
/t/	@	@	*!	*	
/n/	@	*!	*	*	*
/l/	@	*!		*	*
vowels	*!	*		*	*

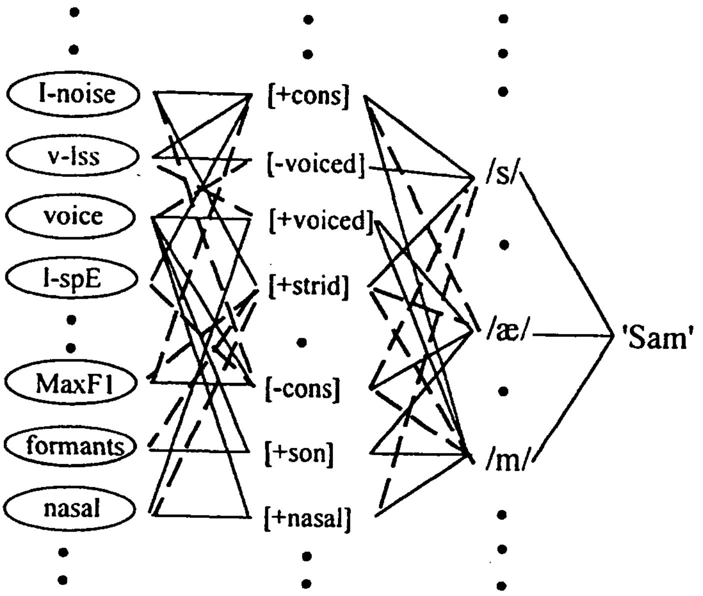
Finally, after the optimal segments (or segment-like units) are chosen, the relevant constraints for lexical retrieval such as phonotactic constraints, syllable structure, obstruent coda unreleasing, etc. will activate to select the optimal sequence of segments to match with the stored lexical items. For example, [k^honmul] or [k^hommul] /k^hosmul/ will be selected over [k^honnul], [k^hosmul], [k^hosul], etc., as illustrated in tableau 8.

Tableau 8

/k^hosmul/	Max-IO	Coda-Neut	N-Assim	CO-Assim	LexMean
@ [k^hommul]	@	@	@	@	@
@ [k^honmul]	@	@	@		@
[k^honnul]	@	@	@	*!	*!
[k^hotmul]	@	@	*!		*
[k^hosmul]	@	*!			*
[k^hopsul]	*!				*
[k^hosul]	*!				*

As seen in tableau 8, [k^hosu] and [k^hosul] will be rejected by violating the constraint *Max-IO* (the number of the input and output segments should be identical). [khosmul] will be disqualified by violating the constraint *Coda-Neut* (obstruents are neutralized in coda position; here /s/ is neutralized to a stop, i.e. [t].). [k^hotmul] will be disqualified through the violation of the constraint *N-Assim* (Stops assimilate to the following nasals to become nasals.). [k^honnul] will be rejected through violation of the constraint *CO-Ass* (the place of the coda regressively and (optionally) assimilates to the one of the onset) and *LexMean* (the meaning of the output should be matched with the lexicon). Hence, [khommul] and [khonmul] will be selected as the optimal words matching with the lexicon, without violating any phonotactic and phonological constraints. In our proposed artificial neural network, this selection process of the optimal ones in the OT framework can be illustrated as in Figure 2.

Figure 2. A section of connectionist neural network.

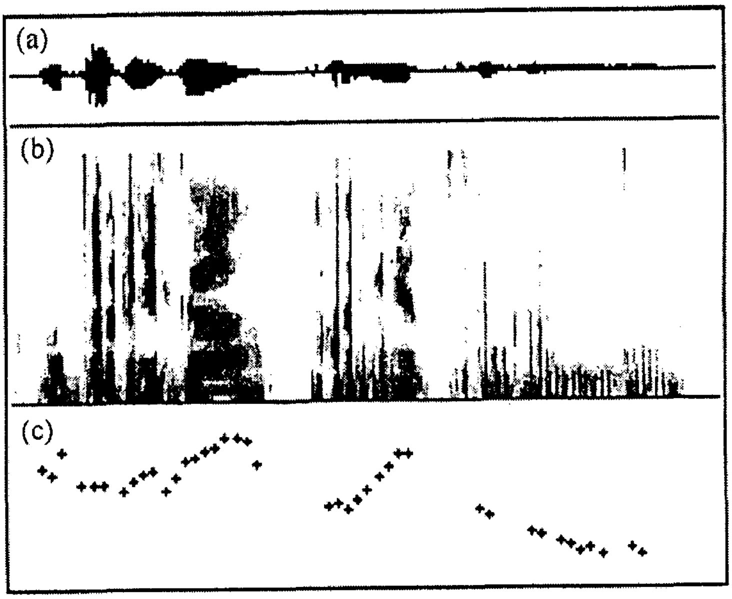


As seen in Figure 2, whenever the acoustic cues are detected in a given speech signal, the corresponding constraints will automatically be fired and activate to send positive (or negative) values to the relevant features, which in turn will activate to choose the optimal segments. Once the optimal segment are selected, by manipulating other relevant phonotactics or phonological rules, the optimal words matching with the lexicon will be chosen.

4. Prosodic features

Recently, Beckman & Edwards (1990), Jun(1993/96), among others, presented the phonetic evidence for the prosodic structures, and showed how phonetic phenomena could be affected by prosodic structures. And Lehiste (1973), Price *et al.* (1991), etc. further showed that the phonetic and prosodic cues can also be used for syntactic disambiguation. Hence by manipulating both the prosodic features such as pitch contour, duration, and the prosody-dependent phonetic cues, we can accomplish more efficient phonological parsing. To see this, consider the following Figure 3, which shows the pitch contour and the spectrogram of 'Hotol-i-ka pa ŋ -e tɨləkanta' (Hotol enters into the room).

Figure 3. The waveform (a), spectrogram (b), and pitch track (c) of /hotolikapa ŋ etɨləkanta/.



As Jun (1993/96) pointed out, Korean standard dialect incorporates the (LH)LH accentual phrase (AP), and stops undergo voicing only within the AP boundary. Furthermore, a pause always follows the intonational phrase(IP). Hence as seen in Figure 3, the sequence of segments /hotolika papa ŋ etɨləkanta/ can be properly dispersed into three IP groups: [hotoliga]$_{IP}$ [pa ŋ -e]$_{IP}$ [tɨləganda]$_{IP}$ by the detection of the rising pitch contour as well as the pause at the end of every IP, which can be illustrated in tableau 9.

Tableau 9.

/hotolikapa ŋ etɨləkanta/	AP	##	VoicingAlloph
@ [hodoriga]$_{IP}$ [pa ŋ e]$_{IP}$ [tɨrəganta]$_{IP}$	@	@	@
[hodori]$_{IP}$ [kapa ŋ e]$_{IP}$ [tɨrəganta]$_{IP}$	*!	*!	*!
[hodori]$_{IP}$ [kapa ŋ]$_{IP}$ [edɨrəganta]$_{IP}$	**!	**!	***!

This three IP(which is AP by default) group disparsing can be also further confirmed by the voicing allophonic cues, as Church (1983/87) and Kim (1990) (among others) showed; that is, the AP-initial stops of /p/ in [paŋ-e]$_{AP}$ and /t/ [tɨləganta]$_{AP}$ show partial aspiration (about 30-70 ms VOT), while the AP-medial stops of /t/ and /k/ in [hotolika]$_{AP}$ undergo voicing to have a negative VOT.

Furthermore, note here that constraints in OT are violable and rankable. Hence, even though there is no detection of pause as in some casual speech, simply detecting the pitch contour and the voicing allophonic cues may suffice to make the correct phonological disparsing into three AP group, as illustrated in tableau 10.

Tableau 10

/hotolikapa ŋ etɨləkanta/	AP	(##)	VoicingAlloph
@ [hodoriga]$_{AP(IP)}$ [pa ŋ e]$_{AP(IP)}$ [tɨrəganta]$_{AP(IP)}$	@	(@)	@
[hodori]$_{AP(IP)}$ [kapa ŋ e]$_{AP(IP)}$ [tɨrəganta]$_{AP(IP)}$	*!	(*)	*!
[hodori]$_{AP(IP)}$ [kapa ŋ]$_{AP(IP)}$ [edɨrəganta]$_{AP(IP)}$	**!	(**)	***!

As seen in tableaux 9 and 10, even though there is no pause as indicated by optional parenthesis (##), the prosodic constraint *AP* contour and the allophonic constraint *Alloph* of voicing are sufficient enough to select the optimal phonological disparsing of [hotoliga]$_{AP(IP)}$ [paŋe]$_{AP(IP)}$ [tɨləganta]$_{AP(IP)}$ over [hotoli]$_{AP(IP)}$ [kapapaŋe]$_{AP(IP)}$ [tɨləganda]$_{AP(IP)}$ and [hotoli]$_{AP(IP)}$ [kapapaŋ]$_{AP(IP)}$ [etiləganda]$_{AP(IP)}$, to make the lexical retrieval more efficient in speech understanding.

5. Summary

So far this paper has shown how multi-acoustic cues can optimally be used in searching for distinctive features, which in turn can be used for selecting the optimal segments. Also, in addition to phonotactics and prosodic structures, by manipulating prosodic features and prosody-dependent phonetic cues, it is shown that more efficient phonological disparsing can be accomplished. Even though certain cues are missing due to various reasons, the optimal search and selection is possible due to the violability and rankability of constraints in the OT framework.

References

Beckman, M. & J. Edwards. 1990. "Lengthening and shorting and the nature of prosodic constituency." in Kingston, J. & M. Beckman (eds) 152-178.

Church, K. 1983/87. *Phonological Parsing in Speech Recognition.*. Boston: Kluwer Academic Pub.

Goldinger, S., D. Pisoni, & P. Luce. 1996. "Speech perception and spoken word recognition." in Lass (1996) ch. 8.

Hardcastle, W. & J. Laver. 1997. *The Handbook of Phonetic Sciences.* Blackwell.

Halle, M. & K. Stevens. 1991. "Knowledge of language and the sounds of speech." in Sundber, Nord, & Carlson eds. *Symposiums on Music, Language, and Brain.* 1-19.

Hayes, Bruce. 1996. "Phonetically-driven phonology: the role of Optimality Theory and inductive grounding." in *Proceedings of the 1996 Milwaukee Conference on Formalism and Functionalism in Linguistics.*

Jakobson, R., Fant, G. & M. Halle. 1952/63. *Preliminaries to Speech Analysis.* MIT Press.

Johnson, Mark. 1994. "Automatic context-sensitive measurement of the acoustic correlates of distinctive features at landmarks." *ICALP94*: 1639-42.

Jun, Sun-Ah. 1993/96. *The Phonetics and Phonology of Korean Prosody.* NY: Garland Publishing.

Kent, R. & C. Read. 1992. *The Acoustic Analysis of Speech.* CA: Singular Publishing Group.

Kent, R., J. Dembowski, & N. Lass. 1996 "The acoustic characteristics of American English." in Lass 1996. ch. 5.

Kim, Kee-Ho. 1990. "Revisiting distinctive feature approach in speech recognition." in *SICONP 91 Proceedings.*.

———. 1991. "The correlation between distinctive feature theories and speech preception." *Journal of English Language and Liturature* (Korean) 37: 783-803.

Kingston, J. and M. Beckman. (eds) 1990. *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech.* Cambridge U Press.

Koenig, W., H. Dunn & L. Lacey. 1946. "The sound spectrograph." *JASA* 18.

Lass, Norman. (ed.) 1996. *Principles of Experimental Phonetics.* Boston: Mosby.

Lehiste, I. 1973. "Phonetic disambiguation of syntactic ambiguity." *Glossa* 7(2): 107-121.

Liberman, A. M., Delattre, P. C., Gerstman, L. J. and Cooper, F. S. 1956. "Tempo of frequency change as a cue for distinguishing classes of speech sounds." *Journal of Experimental Psychology* 52: 127-37

Liu, S. 1995. *Landmark Detection for Distinctive feature-based Speech Recognition.* Ph.D. diss. MIT.

Liu, S. 1996. "Landmark detection for distinctive feature-based speech recognition." *JASA 100*: 3417-3430.

McCarthy, J. & A. Prince. 1995. *Faithfulness and Reduplicative Identity.* Ms. U. of Mass, Amhurst.

Price, P., M. Ostendorf, S. Shattuck-Hufnagel, and C. Fong. 1991. "The use of prosody in syntactic disambiguation." *JASA 90*: 2956-70.

Prince, A. & P. Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar.* Ms. Rutgers U.

Rabiner, L. & B.-H. Juang. 1993. *Fundamentals of Speech Recognition.* NJ: Prentice Hall.

Sagey, E. 1986. *The Representation of Features and Relations in Nonlinear Phonology.* Ph.D. diss. MIT.

Stevens, Ken. 1992. "Lexical access from features. *MIT Speech Communication Group Working Paper* 8: 119-144.

——. (in print) *Acoustic Phonetics.*

▲ 1, 5-Ga, Anam-dong, Sungpuk-ku, Seoul, Korea 136-701
Korea University, College of Liberal Arts,
Department of English Language & Literature 136-701
Tel : +82-2-3290-1988 (O), +82-2-919-5884 (H)
Fax : +82-2-929-1957
e-mail : keehokim@kuccnx.korea.ac.kr