Vowel Fundamental Frequency in Manner Differentiation of Korean Stops and Affricates

Tae-Yeoub Jang*

ABSTRACT

In this study, I investigate the role of post-consonantal fundamental frequency (F0) as a cue for automatic distinction of types of Korean stops and affricates. Rather than examining data obtained by restricting contexts to a minimum to prevent the interference of irrelevant factors, a relatively natural speaker independent speech corpus is analysed. Automatic and statistical approaches are adopted to annotate data, to minimise speaker variability, and to evaluate the results. In spite of possible loss of information during those automatic analyses, statistics obtained suggest that vowel F0 is a useful cue for distinguishing manners of articulation of Korean non-continuant obstruents having the same place of articulation, especially of lax and aspirated stops and affricates. On the basis of the statistics, automatic classification is attempted over the relevant consonants in a specific context where the micro-prosodic effects appear to be maximised. The results confirm the usefulness of this effect in application for Korean phone recognition.

Keywords: microprosody, fundamental frequency, speech recognition

1. INTRODUCTION AND BACKGROUND

A number of previous cross-language phonetic studies have revealed that the fundamental frequency (F0, henceforth) of a vowel is influenced by the type of the neighbouring consonants. Though there are reports on the backward influence of consonant types on the F0 value of the preceding vowel (Kohler 1985), most reports are mainly on the influence of consonants on the following vowel (Gandour 1974, Hombert 1978, Haggard, et al 1970, Ohde 1984, Silverman 1987, among others). A general agreement among those studies can be summarised as: voiceless aspirated consonants raise, while voiced non-aspirated consonants lower, the F0 at the onset of the following vowel.

* Dept. of Linguistics and Centre for Speech Technology Research, University of Edinburgh

As for Korean, which has a three-way phonological, as well as phonetic, distinction of voiceless stop and affricate segments (aspirated, tense and lax)[1], it has been reported that the aspirated and tense stops raise, and lax stops lower the F0 of the following vowel (Kim 1965, Han & Weitzman 1970, Kagaya 1974, Jun 1996). As investigated in Jun (1996), this process of segmental perturbation appears to be especially salient in Korean in comparison to European languages.

Through a series of perception and production tests, Silverman (1987) confirms that the segmental effect can be a significant cue for consonant type differentiation in English. He further shows that the modelling of micro-prosodic[2] perturbation can lead to an improvement of naturalness of the English text-to-speech synthesis system. For example, he generates local F0 perturbations based upon the identity of segments in a syllable and superimposes them to the underlying intonation contour which has been generated separately. Thus combined intonation contour often changes the shape of the original contour considerably, and making the resultant speech sound more natural and acceptable.

Given the consensus on the role of micro-prosody, an inference is made so that the F0 of a vowel can be used in an automatic speech recognition (ASR, henceforth) system to enhance its performance in detecting Korean consonants. To verify this, however, experiments and analyses are required using different data and via different methods from the traditional ones adopted by most previous phonetic studies. In other words, the experimental results mentioned above are not quite appropriate for direct application in practical systems since those experiments are performed on strictly controlled data prepared with the intention of eliminating irrelevant prosodic or segmental effects. Usually tokens containing the target segments were recorded either in a citation form or in a uniform carrier sentence to fix the prosodic environment. The number of subjects was also restricted in order to minimise speaker variability. And then relevant measurements and assessments are done by hand rather than automatically. Though these methods make it possible to minimise variability and errors which most automatic tools are prone to, they cannot be used in the context of

---

1) Following the conventional definition on Korean obstruents, I use the term *tense* to denote such phones as [p', t', k', c'], distinguished from *aspirated* sounds [p^h, t^h, k^h, c^h] as well as *lax* sounds [p, t, k, h]. The three classes are frequently distinguished in terms of laryngeal features: [+ constricted glottis, - stiff vocal chord] for *tense*, [+constricted glottis, + stiff vocal chord] for *aspirated*, and [-constricted glottis, - stiff vocal chord] for *lax* obstruents.

2) The term 'micro-prosody' in this paper refers mainly to local segmental perturbation of F0, apart from other prosodic notions such as duration or intensity, whereas 'macro-prosody' is mainly used to refer to tonal sequence of an utterance which can also be represented in terms of F0.

automatic speech recognition.

In this study, I use a different approach to estimate the influence of the segmental F0. First of all, a relatively large database will be used to verify the F0 characteristics. One of the elements that make the statistical modelling more reliable is apparently a large number of speech tokens. So far as the current state-of-the-art ASR systems are mainly adopting statistical techniques (either Hidden Markov Model or Neural Net, or their hybrid), analyses performed over a large-size data are obviously a useful way of modelling acoustic cues which are directly applicable for such systems. Another advantage of using a large database is the availability of drawing results from various contexts for comparison. Given enough data, this can be done relatively easily saving time and effort which would be spent on careful experimental designs.

Secondly, automatic rather than manual methods of segmentation, annotation and measurement will be adopted. This is inevitable for the ASR application, as when it comes to the test phase of recognising unseen data, an on-the-fly estimation of parameters is supposed to be performed only automatically without any contextual clue provided beforehand.

I will first show that different methods of analysis over different types of data can also produce quite meaningful and apparently more suitable statistical results for practical application. More specifically, the phenomenon that vowel F0 is raised after an aspirated segment and lowered after a lax segment will be shown to be maintained in natural speech data as well. This process is found to be significantly reinforced in a certain context and useful as an additional parameter for the ASR system. The effect of vowel intrinsic F0 values will also be checked to see how they interact with the consonantal effects.

And then, I demonstrate a method of automatic differentiation of consonant types on the basis of the F0 statistics obtained. The classification of the prevocalic aspirated and lax consonants located at the initial syllables of Korean words is performed as a post-processing phase of phone recogniser.

2. DATA

Two different databases are used for different purposes, both of which are provided by *Korean Advanced Institution of Science and Technology* (KAIST) (Choi, *et al.* 1995). One is a continuous speech database on trade negotiation, composed of 3000-word[3] vocabulary, 14746 sentence tokens recorded in a sound-processed room.

3) The exact number varies slightly depending upon the definition of *word*.

It contains voices of 100 male and 50 female speakers from various dialect regions, spoken in a relatively natural fashion. This database is used for developing an automatic phone labeller as described in the next section.

The other database is used for the main purpose of this study, the statistical analysis of the distribution of F0 and its practical application for consonant classification. It consists of isolated word tokens spoken by 34 male and 14 female speakers. It contains 500 Korean city names and the total number of tokens is 7559, of which only 4780 tokens are used for statistical analysis since only those tokens contain at least one stop or affricate segment. Since the proportion of Seoul Korean speakers is not clear from the document which is provided along with the database, some dialect tokens included may have more or less influenced the statistical results. However, I do not exclude any possible non-Seoul dialect tokens assuming the influence should be relatively trivial as all the data tokens are in the form of isolated words.

The data does not contain enough tokens for all twelve (p, p', p^h, t, t', t^h, k, k', k^h, c, c', c^h) target stop and affricate sounds. Especially, tense segments (p', t', k', c') are hard to draw valid statistical results from due to the lack of tokens. The small number of tense consonants, however, does not imply that it is a low-quality database. The frequency of their use in Korean itself is relatively sparse, so the database appears to merely reflect this real situation. It also suggests that differentiation between lax and aspirated stops with better accuracy appears, in practical terms, more useful than the differentiation among all three classes with deteriorated accuracy.

3. AUTOMATIC ANNOTATION

The procedure of auto-labelling is briefly illustrated in Figure 1. First of all, phone-unit hand labelling is performed over 628 sentence tokens. With those annotations, 3-state left-to-right continuous Hidden Markov Models (HMM's) are created for each phone unit using the standard Baum-Welch algorithm. *Mel frequency cepstral coefficients*, *energy*, and their first and second *derivatives* are extracted and trained as parameters. The best string accuracy of this phone recogniser tested over unseen speech data has been 68.29 % with a simple *bigram* language model used to restrict the phone sequence. Based on these HMM's together with sentence orthographic transcription and a letter-to-phone word dictionary, alignments are forced to produce final phone labels for 5736 relevant word tokens. The pronunciation dictionary is devised to take into account morpho-phonemic, phonological, and phonetic

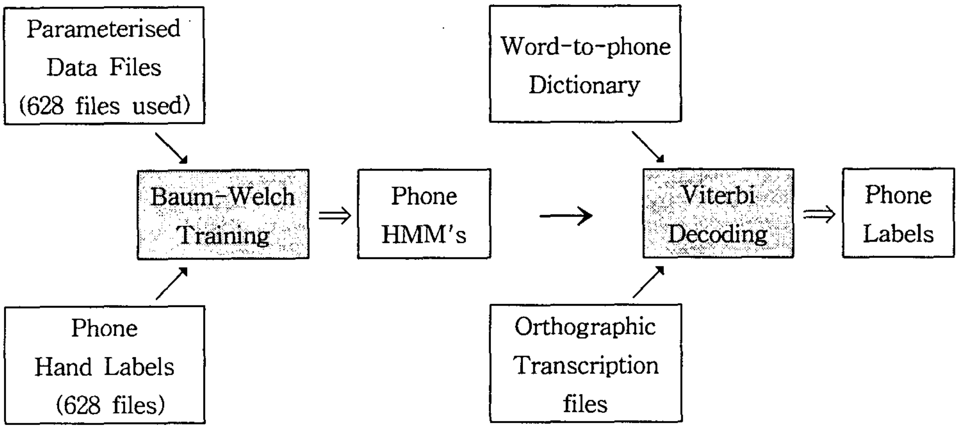

Figure 1. Automatic Labelling Procedure

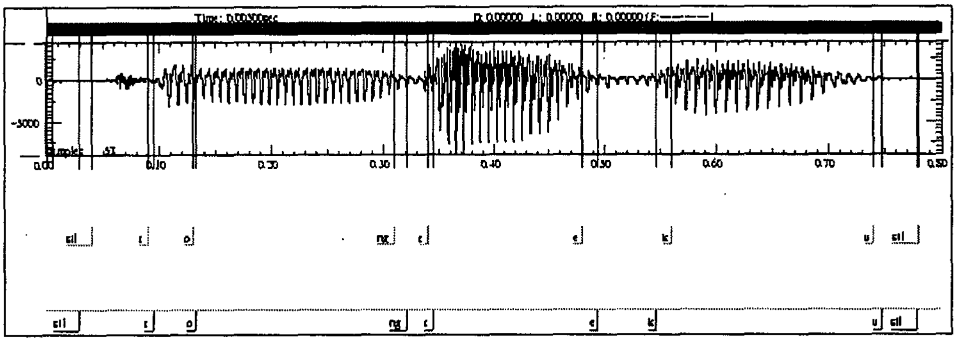

Figure 2. Auto-label Example: The spoken word is *tongteku* [toŋteku] ("east of Taegu city") marked as [sil t o ng t e k u sil]; the bottom line marks are labels created by hand and the upper marks are auto-labels

processes as well as individual and dialectal pronunciation variation. For instance, in case a phone in a word is optionally deletable due to the speaker's speech rate we specify both phone strings with or without the phone in question and let the recogniser make a final decision on the basis of acoustic evidence to pick out the more suitable one for the given speech data. We found this adjustment very useful for obtaining more accurate phone labels for natural speech data in which a great number of processes such as deletions, weakenings, or assimilations take place quite frequently. Figure 2 is an example of auto-label output compared with the corresponding hand label. Misplaced phone boundaries in automatic segmentation might be one of the main reasons preventing more reliable statistical results which are subsequently obtained from those labels. The key to improving the quality of automatic labels is to build more robust bootstrap phone HMM's by providing an increased number of hand labels. This will also improve the statistical results of durational characteristics. However, no hand correction is attempted for demarcation errors so that the type of error itself is consistent when used in an ASR system.

4. F0 EXTRACTION AND NORMALISATION

For F0 estimation, I used *ESPS* (of Entropics) pitch tracking program (named *get_f0*) which is based on *normalised cross correlation function* and *dynamic programming* described in Talkin (1995). Though the performance of this detection algorithm is known to be quite robust it still frequently fails to detect some important regions around the vowel onset position where crucial information for segmental perturbation is contained. Admittedly, this problem is not tackled in this study, which implies that a better F0 estimation algorithm in the future will definitely improve the result of this research.

Considering the number of speakers of both genders, the data must have a large variation of pitch range among the speakers, even among those of the same gender. This variation obviously affects the quality of statistical inferences. To virtually eliminate inter-speaker variability and minimise the inter-gender pitch range difference, each F0 value is readjusted on the basis of the global pitch range whose fixed values of mean and standard deviation of F0 in male speech are calculated over all the voiced vowel tokens in the database.[4] The overall mean value and standard deviation of all vowels are approximately 130 Hz and 25 Hz (Exact values are 132 Hz and 24.52 Hz each). Normalisation is performed to simulate this distribution. For F0 files of each speaker, values are scaled so that two standard deviations worth of values lie between 80 Hz and 180 Hz which indicates again two standard deviations wide either side from the fixed mean 130 Hz.[5] In other words, 95 % of all values are designed to locate between the values 80 and 180 irrespective of the individual speaker's pitch range, assuming each speaker's F0 values are normally distributed.

4) Devoiced vowels which frequently occur following aspirated sounds are not included for calculation of pitch range. In fact, they do not need to be payed attention to as the context itself suggests the preceding consonant be in most cases an aspirated sound.

5) ± 2 SD range is chosen just for instant intuitive interpretation of the normalised F0 values. Varying the range otherwise should not affect the judgment of results regarding the current experiment.

5. STATISTICAL ANALYSIS

5.1 Method

The behaviour of vowel F0 is investigated to see how it varies depending upon the preceding consonant. Measurements are made considering different vowel contexts. They are "after word-initial stop", "at word-final position", and "before word-final coda". The observation of position specific behaviour is quite important in the sense that it can partially reveal the macro-prosodic influence on segmental perturbation. When a word is spoken even in a citation form an utterance level default tune tends to be superimposed on that word unless the speaker deliberately takes strict control over the prosodic structure of the word pronounced. Then it is easy to imagine the possibility of interaction between macro- and micro-prosody. This will be further discussed in the next section.

The effect of vowel intrinsic F0 value is also investigated since this, if significant, will influence the segmental prosody either by intensifying or by attenuating the magnitude. Average normalised values and standard deviations are calculated with a view to further comparisons in each context group, then statistical analyses are performed to verify the significance of variations in F0. As there are more than two manner-of-articulation variables, the *Analysis of Variance (ANOVA) Test* is performed instead of *t-test* which is known to be less accurate for three or more variables by increasing Type I error. As the ANOVA test only determines whether any of three variables is statistically significant against any other variable without saying anything about pairwise relation, a supplementary test, what is called *Tukey's multiple comparisons procedure* (Devore & Farnum 1999:395), is subsequently performed, which takes into account all possible pairwise comparisons. Both tests are performed at the significance level of 0.01 (or 99 % confidence level).

5.2 Results and Discussion

The calculated values of vowel F0, varying depending upon each preceding consonant class, are summarised in Table 1. Three unit cells of each context indicate from top to bottom: the average normalised F0 value in Hz, standard deviation, and number of tokens. For example, the F0 of vowels after a lax stop at a word final open syllable (Table 1. (c)(i)) has, as its distribution, mean value 118.05 Hz and standard deviation, 11.14, calculated from 842 measured tokens. The results of statistical significance test are summarised in Table 2.

The general tendency of vowel F0 being higher after aspirated consonants than after lax consonants is widely shown regardless of contexts. Note this is also statistically verified: each result of significant tests in Table 2 shows that vowel F0 values are differentiated depending on whether the preceding segment is an aspirated or a lax consonant.

Table 1. Normalised F0 Values: Each row represents, from up to bottom, *mean F0*, *standard deviation*, and *number of tokens*.

	Context	(i) After Lax	(ii) After Tense	(iii) After Aspirated
(a)	Overall Average	133.25	133.91	147.94
		13.97	20.65	22.69
		5518	193	2122
(b)	# C _	137.05	-	172.72
		11.18	-	16.65
		2896	0	575
(c)	C _ #	118.05	116.46	123.15
		11.14	11.81	12.06
		842	50	290
(d)	C _ C #	130.21	132.34	138.41
		11.84	15.46	12.00
		1410	106	1081

Table 2. Statistical Significance: Results of one-way ANOVA and Tukey's multiple comparisons procedure

Context	ANOVA	Tukey's Pairwise Comparison		
		Test Pair	Mean Distance Compared with Threshold	Significance
Overall Average	F = 578.32, p < 0.001	lax-tns	0.66 < 3.62	
		lax-asp	14.69 > 1.36	✓
		asp-tns	14.02 > 3.71	✓
# C _	F = 4060.4, p < 0.001	lax-tns	NA	
		lax-asp	35.67 > 1.44	✓
		asp-tns	NA	
C _ #	F = 23.10, p < 0.001	lax-tns	1.58 < 4.84	
		lax-asp	5.11 > 2.26	✓
		asp-tns	6.69 > 5.09	✓
C _ C #	F = 141.41, p < 0.001	lax-tns	2.13 < 3.54	
		lax-asp	8.19 > 1.42	✓
		asp-tns	6.06 > 3.58	✓

The F0 value of vowels preceded by tense consonants are not as high as shown in past phonetic experiments (Han & Weitzman 1970:117, Kagaya 1974:169, among others). This result, however, is not yet meaningful enough to argue against those results of phonetic researches on the F0 raising effect of tense obstruents, because the present result on this issue is not sufficiently reliable due to the lack of tense data tokens. After all, as the major concern in this study is the differentiation between aspirated and lax consonants, the role of tense consonants is not seriously counted for the moment. Even if it is confirmed by other experiments with enough data that the segmental F0 perturbation is not quite useful for distinguishing tense sounds, that does not necessarily debase the value of current study as a whole, since there seems to be other cues that differentiate tense sounds. For example, phonetic studies have found the usefulness of *voice onset time* (VOT) and/or *closure duration* in differentiating stop consonants (Yun & Jang 1999). In one of my and my colleague's ongoing pilot experiments, it is likely that those temporal charactestics are especially useful for Korean tense stops rather than the other two types. It implies that more effective classification in practical speech systems will be possible if we take advantage of those two complementary prosodic cues simultaneously.

The degrees of F0 difference between vowels at post-lax positions and post-aspirated positions are not consistent through the locations of the phones in question. The difference is most conspicuous when an obstruent is located word initially (Table 1, (b)(i) and (b)(iii)) compared with the difference in other environments, which is visualised by the normalised F0 distribution graphs in Figure 3 and 4.

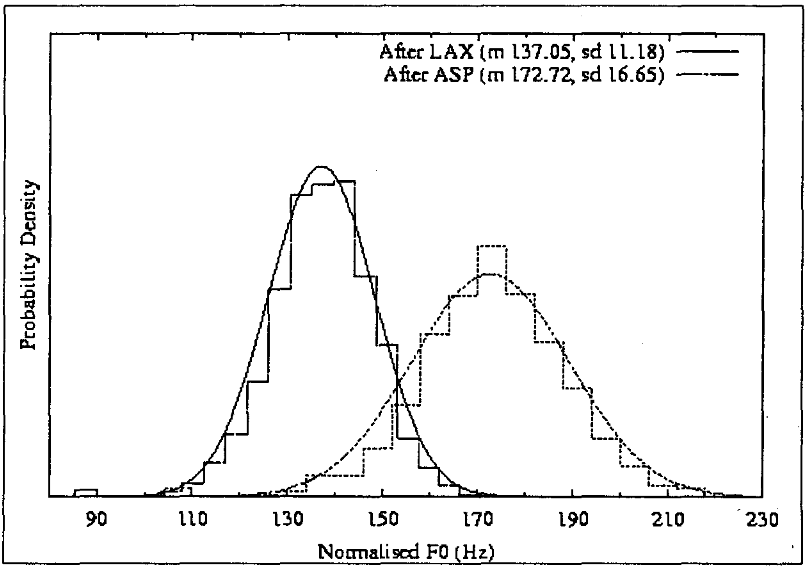

Figure 3. Histograms and Normalised Curves of Post-consonantal F0 Distributions (F0 of Vowels in the First Syllable)

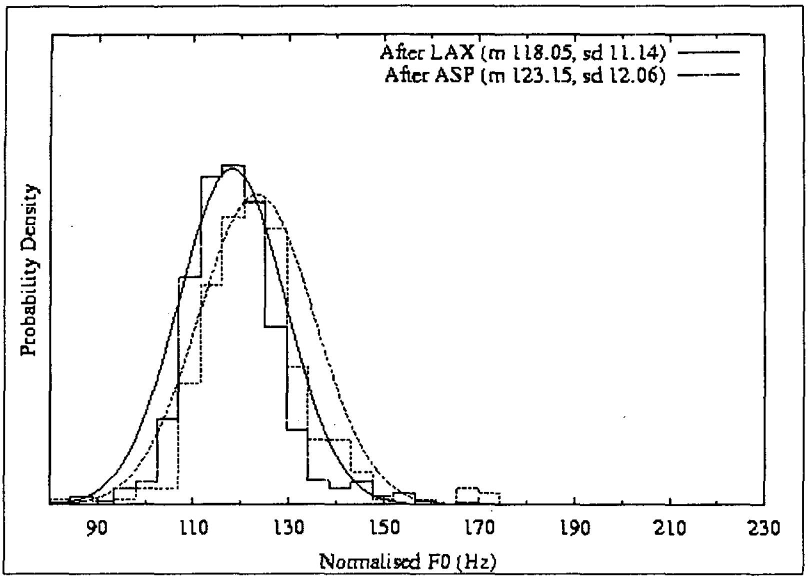

Figure 4. Histograms and Normalised Curves of Post-consonantal F0 Distributions (F0 of Vowels in the Last Syllable)

The explanation for this different result can be sought in conjunction with the macro-prosodic influence. As has been briefly mentioned in the previous section, each word token is usually produced under the influence of global utterance intonational patterns as well as local segmental perturbations. It is reasonable to think that the general tune of declarative sentences is superimposed on the segmental structure of each word, since every production of an isolated word token, however small its size is, with only an identifiable duration of pause in between, can be classified as a declarative sentence utterance in terms of intonation analysis. This leads to an inference that an *Intonational Phrase* (IP, henceforth) with a falling boundary tone is a default prosodic structure of isolated word production.

Getting back to the inconsistent result under consideration, the last syllable of the word tokens can now be said to bear the IP boundary tone, and consequently, the relevant segmental perturbation has been considerably preempted by that macro-prosody. Though this might be the case, note that there still exist some discernible, and statistically verified differences between aspirated stop and lax stop classes in Table 1 (c)(d). It might be evidence of micro-prosodic effects being preserved even after macro-prosodic influence, which may be an encouraging result for later implementation into the practical system.

The result of the relatively prominent difference at syllable initial position (Table 1(b)) appears to be directly associated with the characteristics of another intonationally defined prosodic unit called *Accentual Phrase* (AP, henceforth) of Korean. Jun (1998) describes AP as a prosodic unit which is one level lower than IP and higher than *Phonological Word*, observing *Strict Layer Hypothesis* (Selkirk 1984) in the hierarchical structure of Korean prosody. The current result agrees well with her suggestion that the AP initial H tone of Seoul Korean is accounted for in terms of segmental perturbation.

I also evaluated whether there is a so-called *vowel intrinsic F0* effect and the extent to which it interacts with the stop manner influences. This is necessary for reliable modelling of F0 perturbation at stop-vowel sequences. The general agreement on this effect is that high vowels have higher intrinsic F0 than low vowels. The result shown in Table 3 seems to retain this tendency, but the size of differences between values of high and low vowel F0 is relatively small. In brief, the vowel intrinsic F0 factor does not seem to be as salient in Korean as in other languages like English and German (Ladd & Silverman 1984, Silverman 1987). As Korean, unlike English, does not have the *tense/lax* contrast of vowels in the phonological level, no two words are made as a minimal pair due only to the difference of vowels in the same position of each word.[6] The lack of such contrast is likely to make it unnecessary for speakers to pronounce high vowels with more prominence or with a distinctively higher F0. If this is the case, the vowel intrinsic pitch must be a less

Table 3. Vowel Intrinsic F0 Effect: 3 rows of each context represent, from up to bottom, mean F0 (Hz), standard deviation, and number of tokens.

Syllable Position	Preceding Stop	High Vowel	Mid Vowel	Low Vowel
Word Initial	Lax	140.54	136.66	131.92
		12.56	9.55	9.68
		1083	1096	509
	Tense	No Tokens		
	Aspirated	175.64	174.52	168.60
		17.86	17.66	14.32
		101	264	119
Word Final	Lax	120.96	128.58	128.12
		12.66	12.92	11.40
		739	1038	247
	Tense	Not enough tokens		
	Aspirated	139.00	133.26	136.23
		21.92	13.03	11.17
		62	904	161

6) English Examples of tense/lax contrast are: seat/sit, gate/get, two/to.

influential element of F0 fluctuation, and the pre-consonantal effect becomes relatively more important, which is another positive indication for the purpose of current study. However, further scrutiny with more refined experimental design is required to draw a firm conclusion on this issue.

There is a phonological process by which a vowel is devoiced in the vicinity of voiceless obstruents. Questions may arise regarding this devoicing process for fear that it should make the vowel F0 extraction impossible. However, in the data of this study only a small number of vowels (13 cases) are totally devoiced to the extent that its mean F0 value cannot be extracted.

6. AUTOMATIC DIFFERENTIATION

It is not always the case that the verification of statistical significance implies that it can be exploited in an implementation. In order to see the segmental F0 effect is also useful for automatic speech recognition, an experiment of automatic classification between lax and aspirated consonants is attempted for the data tokens placed at the most prominent positions: close to the left edge of an intonationally defined prosodic boundary.

6.1 Method

Two sets of phone HMM's are used to perform two separate versions of automatic recognition. For both the tests, simple phone-*bigram* language models are generated from the label files of each database. The same isolated word data is used for the test as in statistical analysis. But this time a smaller number of word tokens (1377) are selected, namely those tokens beginning with one of 8 lax or aspirated non-continuant consonants (p, t, k, c, p^h, t^h, k^h, c^h) followed by a vowel.

In the first version of recognition test, the same HMM set is used as in the automatic labelling phase mentioned in Section 3. This HMM set, composed of 43 phone models, is used for generating baseline results without any consideration of the information on post-obstruent vowel F0. The recognition results obtained from those ordinary models are for the comparison with the results from the micro-prosody differentiation model.

The other version of recognition test employs vowel F0. This consists of a few steps. First, phone-level recognition via standard *viterbi decoding* is performed. The HMM's used here are the same as those of the baseline system except that the models of stops and affricates do not have any manner distinction. In other words, only place-separated symbols are forced to output in the form of C, P, T, K, each representing voiceless affricate, bilabial stop, dental stop, and velar stop. In parallel, an F0 file for each test token is generated and then normalisation is performed. F0 generation is the same as described in Section 4, but normalisation cannot be done in the same way as before since it is no longer the case that a large number of same-speaker data tokens are provided all at once in the test phase. Therefore, only utterance-level normalisation is performed. That is to say, all 10msec-frame-level F0 values of voiced sounds in an utterance are collected and then normalisation is performed as stated in Section 4. Some information may be lost during this step because the data tokens (ie. isolated words) are not long enough to contain as much voiced period as is required for a reasonable averaging. This difficulty will be alleviated when the utterance-level F0 normalisation is done over normal sentence tokens rather than isolated words as in the current experiment. The likelihood of thus obtained post-consonantal vowel F0 value in each class is calculated in terms of the class conditional probability multiplied by the prior probability of the consonant class:

$$P(Class|F0) \simeq P(F0|Class)\ P(Class)$$

The best scale of prior probability is found through the heuristic adjustment. Finally, the decision on the manner of the preceding consonant is made on the basis of the posterior probability.

6.2 Results

Table 3 and 4 summarise the compared classification results obtained from the two models. The third column *Manner Decision Correctness* shows how correctly the manner distinction is made regardless of the other substitutions, insertions, and deletions of phones. Overall measures of each table indicate better performance of the model using micro-prosodic information.

In the current micro-prosody model, the *aspirated/lax* distinction is made solely in terms of the post-consonantal F0 and no other cues are taken into account. This is to concentrate the experiment on the purpose of checking the role of the micro-prosodic effect. Thus, if the F0 effect is devised to cooperate with or assist the performance of other standard parameters of recognition, better phone accuracy is expected to be obtained.

Table 3: Classification Results from Baseline Model

	Phone Accuraccy (%)	Manner Decision Correctness (%)
Lax	23.29	79.86
Aspirated	21.69	64.19
Overall	**22.95**	**69.01**

Table 4: Classification Results from F0 Model

	Phone Accuraccy (%)	Manner Decision Correctness (%)
Lax	29.67	92.16
Aspirated	23.05	76.95
Overall	**28.25**	**87.69**

7. CONCLUSIONS

Automatic statistical methods of F0 analysis are found to be useful in capturing the pattern of vowel F0 fluctuation influenced by the manner of articulation of the preceding consonant in Korean. It turns out that F0 of a vowel following an aspirated stop or affricate is significantly higher than when it is following a corresponding lax counterpart. The effect is particularly boosted at the initial syllable of each word token while it wanes at the last syllable.

As for the automatic classification of *lax/aspirated* stops and affricates in word-initial syllables, the micro-prosody model only using F0 information of the following vowel shows better performance than the baseline model created from the standard HMM technique using various other conventional parameters. This leads to an expectation that through combination with other standard parameters, micro-prosodic information can improve the quality of ASR systems.

REFERENCES

Choi, I. J., J. R. Park, D. Y. Kim, H. Y. Jeong, & C. K. Un. 1995. "Speech Data Collection for Korean Speech Recognition." *The Journal of the Acoustic Society of Korea, 14(4),* 74-81. (in Korean).

Devore, J & N. Farnum. 1999. *Applied Statistics for Engineers and Scientists.* Duxbury Press.

Gandor, J. T. 1974. "Consonant types and tone in Siamese." *Journal of Phonetics, 2,* 337-350.

Haggard, M., S. Ambler & M. Callow. 1970. "Pitch as a Voicing Cue." *Journal of Acoustic Society of America, 47,* 613-617.

Han, M. & R. Weitzman. 1970. "Acoustic features of Korean /P, T, K/, /p, t, k/, and /p^h, t^h, k^h/." *Phonetica, 22,* 112-128.

Hombert, J.-M. 1978. "Consonant Types, Vowel Quality, and Tone." In Fromkin, V. (Eds.), *Tone: A linguistic survey,* 77-111.

Jun, S.-A. 1993. *The Phonetics and Phonology of Korean Prosody.* PhD Dissertation. The Ohio State University.

Jun, S.-A. To appear. "The Accentual Phrase in the Korean prosodic hierarchy." *Phonology Yearbook 15(2).*

Kagaya, R. 1974. "A Fiberscopic and Acoustic Study of the Korean Stops, Affricates, and Fricatives." *Journal of Phonetics, 2,* 161-180.

Kim, C.-W. 1965. "On the autonomy of the tensity feature in stop classification (with special reference to Korean tops)." *Word, 21,* 339-359.

Kohler, J. K. 1985. "F0 in the perception of lenis and fortis plosives." *Journal of Acoustic Society of America, 48(1),* 21-32.

Ladd, D. R. & K. E. A. Silverman. 1984. "Vowel intrinsic pitch in conneted speech." *Phonetica,* 41, 31-40.

Ohde, R. 1984. "Fundamental frequency as an acoustic correlates of stop consonant voicing." *Journal of Acoustic Society of America, 75,* 224-230.

Selkirk, E. 1984. *Phonology and Syntax: The Relation between Sound and Structure.* MIT Press.

Silverman, K. E. A. 1987. *The structure and processing of fundamental frequency contour.* Ph.D. Dissertation. University of Cambridge.

Talkin, D. 1995. "A Robust Algorithm for Pitch Tracking (RAPT)." In Kleijn, W. B. and K. K. Paliwal (Eds.), *Speech Coding and Synthesis.* New York: Elsevier.

Yun, W. & T. Jang. 1999. "Statistical analysis of Korean stop durations and its application for speech recognition." In *International Conference on Speech Processing '99,* 1, 305-310.

Received : December 12, 1999

Accepted : February 20, 2000

▲ Centre for Speech Technology Research
University of Edinburgh
80 South Bridge
Edinburgh EH1 1HN, UK
Tel: +44-131-667-2797
e-mail: tyjang@cstr.ed.ac.uk