Shapes of Vowel F0 Contours Influenced by Preceding Obstruents of Different Types*

-Automatic Analyses Using Tilt Parameters-

Tae-Yeoub Jang**

ABSTRACT

The fundamental frequency of a vowel is known to be affected by the identity of the preceding consonant. The general agreement is that strong consonants trigger higher F0 than weak consonants. However, there has been a disagreement on the shape of this segmentally affected F0 contours. Some studies report that shapes of contours are differentiated based on the consonant type, but others regard this observation as misleading. This research attempts to resolve this controversy by investigating shapes and slopes of F0 contours of Korean word level speech data produced by four male speakers. Instead of entirely relying on traditional human intuition and judgment, I employed an automatic F0 contour analysis technique known as *tilt* parameterisation (Taylor 2000). After necessary manipulation of an F0 contour of each data token, various parameters are collapsed into a single tilt value which directly indicates the shape of the contour. The result, in terms of statistical inference, shows that it is not viable to conclude that the type of consonant is significantly related to the shape of F0 contour. A supplementary measurement is also made to see if the slope of each contour bears meaningful information. Unlike shapes themselves, slopes are suspected to be practically more practical for consonantal differentiation, although confirmation is required through further refined experiments.

Keywords: F0 contour, microprosody, segmental prosody, tilt, automatic F0 analysis, fundamental frequency, F0 contour shape

1. Introduction

It has been confirmed in a number of studies that segmental identity affect formulation of F0 of the following vowel. The conditioning sound classes causing this effect, which I will call *segmental F0* (SF0, henceforth) effect, can be different language specifically according to the phonetic or phonemic inventory of each language. For instance, voiceless

* This work was supported by Hankuk University of Foreign Studies Research Fund of 2003.
** Department of English, Hankuk University of Foreign Studies

consonants cause higher F0 while voiced consonants cause lower F0 of the following vowel in English (House & Fairbanks 1953, Lehiste & Peterson 1961, Haggard *et al.* 1970, Mohr 1971, Löfqvist 1975, Hombert 1978, Umeda 1981, Ohde 1984, Silverman 1986, Terken 1995) or in German (Kohler 1982). On the contrary, aspirated and fortis consonants trigger higher F0 while lenis consonants trigger lower F0 at the same position in Korean (Kim 1965, Han & Weitzman 1967, 1970, Hardcastle 1973, Kagaya 1974, Jun 1996, Kim *et al.* 2002).

Although the range of SF0, in terms of magnitude, has been agreed upon by most studies, forms of local SF0 contours have been controversial. On one hand, investigators report that there is a 'rise-fall dichotomy'. That is, F0 contours keep falling from the vowel onset after voiceless consonants, whereas they rise for a certain period before starting to fall after voiced consonants (Lehiste & Peterson 1961, Haggard *et al.* 1970, Gandour 1974, Hombert 1978, Lea 1980). On the other hand, others argue that there is no such dichotomy of shape or direction and the post-stop contours basically move downward regardless of the class of the preceding consonant (Kohler 1982, Ohde 1984, Silverman 1986, Silverman 1987). They state that the seemingly apparent rise-fall dichotomy can be attributed to a failure to properly control or neutralise other intervening prosodic factors. Indirect support for no-dichotomy is found in other experiments, although these were not directly designed to investigate the issue in question. In the studies of Umeda (1981) and Löfqvist (1975), it is shown that F0 after voiceless stops can rise for a short duration depending on the location of prosodic features such as lexical stress, sentential stress, focusing, or intonational pitch accents.

There are similar contradictory reports of experiments with Korean data. For example, Han & Weitzman (1970) state that F0 after lenis stops begins at a relatively low level and then rises to a relative peak after 50-100 msec while F0 after fortis or aspirated stops begins from a point with a high value and either stays constant or starts to fall within the same time span (page 116). On the contrary, Kim (1968; cited in Mohr (1971)) found that F0 contours which follow three classes always fall, though the slope becomes steepest when preceded by aspirated stops and least steep when preceded by lenis stops. Jun (1996) also demonstrates that F0 values are consistently falling after all three types of consonants.

In this paper I dtailed here in attempt to investigate this disagreement. The special design of the current experiment as compared to previous studies can be summarised as follows. First, a relatively great number of data tokens are collected. As it is possible that the disagreement among previous studies is due to a lack of data tokens, more sufficient data are necessary. The number of participants in the current study cannot be said to be considerably great compared with other studies, but I attempted to increase the number of tokens by taking into account various phonetic contexts of relevant segments. Second, automatic data manipulation and measurement techniques are employed. As the future goal

of this research is its implementation into a practical speech technology system an automatised process is essential. Even though it is inevitable that automatisation of the entire process will cause some degrees of information loss, I expect continued tuning and iterated verification of the results are expected to improve the credibility of the automatic techniques.

Details of the experimental design, processing, and results are described from the next section.

2. Data

2.1 Data Creation

I used the data created for investigating magnitudes of segmental F0 effect described in Jang (2000b). Below is a summary of how the corpus is generated.

A corpus is made to consist of words beginning with Korean stops and affricates. For recording, a list of 216 two-syllable isolated words, selected from a Korean dictionary, was prepared. Most words include one of the 12 non-continuant obstruents of Korean at the initial position; but those consonants of three different types are not evenly distributed in the words. In Korean, the number of words beginning with a lenis obstruent is much greater than the number of the words with the other two types at the same position. Thus the list extracted for the current study contains more words with lenis stop sounds at the word-initial position. However, an efforts has been made to include as many words with fortis and aspirated sounds as possible. This is for the purpose of valid comparison among obstruent classes at the experiment.

Unlike most phonetic experiments, the vowel type of each syllable was not artificially fixed and vowels of various different heights were as evenly distributed as possible.[1)]

All the tokens were spoken within a fixed carrier sentence in the form of:

(1) *i-keos-eun* _____ *cheo-leom po-in-ta*
this+TOP like look+FIN
'This looks like _____'

1) Note that the vowel intrinsic F0 (i.e., higher vowels causing higher F0) can affect the shape of the F0 contour undermining the result of the current study, but Jang (2000b: 89) has already shown that its effect is relatively small so that it does not preempt the segmental F0 effect. Thus, I assume that the intrinsic vowel F0 effect does not seriously affect the shape of the F0 contour considerably.

This uniform design is intended to keep the macroprosodic effects under control This procedure is necessary as non-uniorm intervention of the underlying macroprosodic structure will undermine reliability of the statistical estimation of the segmental effects, if not totally obliterate them.

Four male speakers participated in recording: JSH, KHK, TSS, and WHY. All of these men were brought up and educated in the area where Seoul Korean is spoken, and three of them were at the same time graduate students majoring in subjects other than phonetics or linguistics. Only WHY is a student of phonetics, and aware of the purpose of the experiment but I assume the influence of his technical knowledge is not large enough to alter the results to a considerable degree. Sentences were recorded in randomised order. Recorded data are quantised to a 16 bit memory level with a 16 KHz sampling rate. Therefore, a total of 4320 tokens (216 words×4 speakers×5 iterations) were obtained.

2.2 Phonetic Annotation

As the orthographic information has already been provided for each data token and the F0 track can be produced independent of segmental level, it may not be necessary to have phonetic segmental annotations for the current study. However, it is also the case that, with the information of vowel boundaries, a more accurate location of the relevant F0 contour can be obtained with less effort and that processing time can be saved considerably. Thus, automatic labelling is performed using standard automatic speech recognition techniques. Detailed methods of this process are described in Jang (2000b: 78).

3. Experimental Methods

3.1 F0 extraction and Normalisation

For F0 estimation, I used the *get_f0* pitch tracking program of *ESPS* (Entropic 1998) which is based on the algorithms known as *normalised cross correlation* and *dynamic programming* as described in Talkin (1995). Though the performance of this detection algorithm is known to be quite robust the program still frequently fails to detect some important regions around the vowel onset position where crucial information for segmental perturbation is contained. This problem is not tackled in this study and a better F0 estimation algorithm in the future is expected to further improve the result of this research.

It sometimes happened that F0 of some vowels was not captured due to various

reasons. For example, when a high vowel is surrounded by obstruents it frequently gets shortened, devoiced or even deleted. In those cases, it is not possible to extract F0 values stable enough for investigation. Those vowels were excluded in measurement lest they unduly influence the measurement results. Other tokens that are judged to be unsuitable for F0 analysis are not taken into account, either. Consequently, the number of tokens shown in the result below, which is 3318, is quite less than the number of spoken tokens, which is 4320.

The pitch range of the four speakers cannot be assumed to be the same. Moreover, within-speaker range may not be consistent when an utterance is pronounced several times. In general, such variation critically affect the quality of statistical inferences during the experiments in which investigation of F0 magnitudes is a major concern. Although the current experiment does not probe into such F0 magnitudes, different pitch range of speakers can also trigger different F0 shapes, which will undermine the reliability of the current analysis. Consequently, tackling this problem is necessary.

To virtually eliminate inter-speaker variability each F0 value — and subsequently F0 contour — is readjusted on the basis of the adult male speakers' global pitch range whose fixed values are mean, 130 Hz, and standard deviation, 25 Hz, respectively. These values are borrowed from Jang (2000b), who obtained them through a F0 analysis on a relatively large speech corpus originally built for automatic speech recognition research (see page 123). 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 two standard deviations either side of the fixed mean 130 Hz. In other words, 95% out of all values are made to fall between the values 80 and 180 irrespective of the individual speaker's pitch range, assuming that each speaker's F0 values are normally distributed. The algorithm used for this process is shown as follows:

(2) Formula for speaker normalisation of F0 range

$$F0[i]_{normalised} = \left(\left(\frac{F0[i] - \mu_s}{4\sigma_s} + 0.5 \right) (High - Low) \right) + Low$$

where μs and σs stand for the mean and standard deviation of F0 value for each utterance and Low and High specify the designated pitch range. As is mentioned, the values 80 and 180 are allocated. F0[i] means that F0 of i'th sample point.

In brief, all the F0 values are redistributed based on a fixed mean value 130 Hz, which is found to be the average vowel F0 of all male speakers.[2)]

3.2 Tilt Paramaterization

To analyse the shape of the F0 contours I used the *Tilt intonation model* (Taylor & Black 1994, Taylor 2000). An advantage of the Tilt model is that its parameters can be extracted automatically.

As Figure 1 illustrates, the Tilt model characterises the relative shape of an F0 contour by a single number which ranges from -1 to 1. As the shape of a contour gets closer to a pure rise the Tilt value gets near 1, and pure fall near -1.

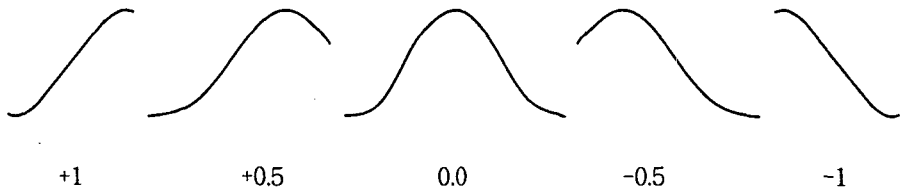

Figure 1. Examples of F0 contours and their corresponding Tilt values

When a F0 contour is given in the form of a series of numbers, as is always the case in digital processing of speech signals, a Tilt value representing the contour shape is automatically extracted. For this purpose, four parameters are employed: *Rise Duration, Fall Duration, Rise Amplitude, Fall Amplitude,* which are schematised in Figure 2.

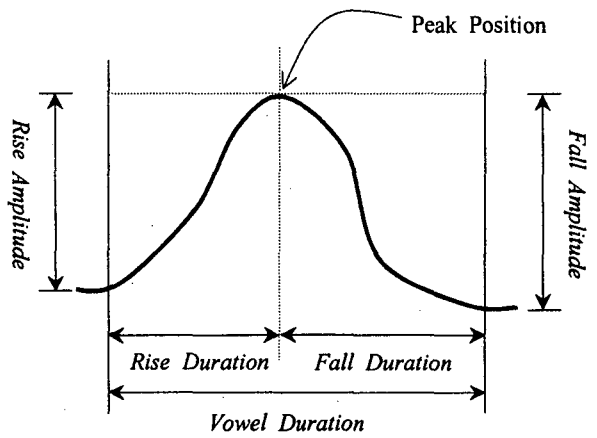

Figure 2. Elements of tilt parameter calculation

2) For the z-score normalisation, as in the present experiment, to be effective, normality in distribution of data needs to be presupposed. Jang (2000a) confirms that segmental F0 values are normally distributed at least in Korean.

Once the individual tilt parameters are calculated, they are collapsed together to produce a final Tilt value which finally defines the general shape of the final contour. This procedure is done in terms of the formula in (3).

(3) Formula for Tilt value generation

$$tilt = \frac{1}{2}\left(\frac{|A_{rise}| - |A_{fall}|}{|A_{rise}| + |A_{fall}|} + \frac{D_{rise} - D_{fall}}{D_{rise} + D_{fall}} \right)$$

Where A denotes *Amplitude* and D denotes *Duration*.

3.3 Measurements

Tilt parameters of F0 contours of each vowel after stops and affricates in word initial syllables was calculated. Programs are written for automatic generation of Tilt values for all the F0 contours generated as explained previously. Then the statistical significance was calculated in terms of a two-tailed single factor ANOVA test followed by subsequent Tukey's pairwise judgment.

4. Results and Discussion

The results of Tilt analysis are in the form of decimal values between -1 and 1. These values directly reflect the shape of each contour. Based on these values the effects of segmental F0 formulated by the three different consonantal types, fortis, lenis, and aspirated, are compared.

The primary comparison is provided in terms of F0 shapes as Tilt values. Then a supplementary analysis is conducted to determine whether the slope of each contour has any significant bearing on consonantal characteristics.

4.1 F0 shapes

Table 1 shows the overall result of Tilt values for each contour.

Table 1. Average tilt values for each type of preceding consonants

	Mean	St. Dv.	Num	% of Contours Having Negative Value
Lenis	-0.74	0.56	1749	88.56
Fortis	-0.68	0.58	618	88.19
Aspirated	-0.68	0.49	944	90.25
			3,311	

Each mean value is obtained via averaging all the individual values for each category. The last column of the table reveals how many negative values, or fall-like, values are produced as compared to other shapes.

That the mean values for all three types lie between -0.5 and -1 implies that the slope of contours tends to be consistently falling (i.e., similar to the 5th or 6th slope illustrated in Figure 1 above), irrespective of the type of the preceeding consonant. This result obviously conflicts with the results shown in Han & Weitzman (1970). To match with their observation, strong (fortis and aspirated) consonants are supposed to induce tilt values closer to -1 than weak (lenis) ones, which is not the case in the current study.

The rate of negative value is a little higher when preceded by aspirated sounds (90.25%). Nevertheless, the extent of the difference between other classes is too small to conclude that the aspirated stops cause the F0 contour to fall from a higher peak position.

The significance test summarised in Table 2 further confirms the inappropriateness in differentiating consonant classes based on the shapes of F0 contour.

Table 2. Summary of the Significance Test

Summary of ANOVA				
Source	df	SS	MS	F
Factor	2	2.84	1.420	4.83
Error	3308	973.31	0.294	
Total	3310	976.15		
Result	$p > 0.01$ (Non-significant)			

Consequently, it can be concluded that it is not appropriate to employ the shape of the contour in differentiating consonant types in any kind of spoken language systems.

It is not clear at the moment what made the measurements reported in Han & Weitzman (1970) clash with the current result. One possibility is that their investigation is based on the data produced by only two (one male and one female) informants and their specific characteristics may have disguised the general tendency of segmental F0 contour shaping. If this possibility is the case, it can be inferred that the shape of segmental F0

contour is basically a part of personal voice characteristics and consequently cannot be used as an invariable acoustic correlate of consonant manners.

A closer examination of the current experimental results further supports that conclusion. As shown in Table 3, no consistency in tendency is found among four speakers. Especially, the result of the speakers, KHK and TSS, shows that F0 after aspirated consonants reachts its lowest point, which is striking and contrary to expectation.

Table 3. Consonant-type specific average tilt values for each speaker

Speaker	C Type	Mean	St. Dv.	Num
JSH	Lenis	-0.72	0.60	463
	Fortis	-0.65	0.60	155
	Aspirated	-0.78	0.40	249
KHK	Lenis	-0.78	0.51	443
	Fortis	-0.65	0.60	155
	Aspirated	-0.59	0.50	230
TSS	Lenis	-0.62	0.62	397
	Fortis	-0.62	0.60	154
	Aspirated	-0.48	0.61	233
WHY	Lenis	-0.81	0.49	446
	Fortis	-0.78	0.48	154
	Aspirated	-0.84	0.31	232
Total				3,311

Another possible reason for varying results is the different experimental methods between the previous and this one. As has been described, the current experiment is designed to employ only automatic measurement techniques. Although this approach has an obvious advantage in the implementation of speech systems, unavoidable processing errors, whether caused by imperfect algorithms or erroneous application, may have caused some significant the reliability of information to be permanently lost. Admittedly, it is still premature to determine how reliable the automatic techniques used in the current research are and how the results they tend to produce differ from the manual, intuitive methods used in most previous studies.

4.2 Slope Analyses

Although all three types of obstruents turned out to be consistent in giving rise to nagative F0 contours of the following vowel, the rate of fall cannot be assumed to be uniform. To clarify this rate, I calculated the slope of each contour using linear regression.

As the peak position is supposed to be located automatically, the declining slope can be easily discovered mathematically.

The result of slope calculation reveals: -3.27 after lenis, -2.29 after fortis, and -4.14 after aspirated sounds. The difference between each pair turned out to be highly significant ($F(2, 3308)=48.23$, $p<0.001$).

The steeper slope of post-aspirated F0 contours might be reasonable considering the segmental F0 effect which implies that higher F0 occurs after aspirated sounds. However, the least steep slope that occur after tense sounds is then initially surprising since the fortis sounds are also known to be included in the strong consonant category causing higher F0 than lenis sounds. The explanation for this phenomenon can be found in the difference in vowel duration which affects the slope of the contour. That is, the duration of vowels after fortis sounds (78.77 msec) was found to be considerably longer than after lenis (69.91 msec) or aspirated sounds (59.35 msec).

The results suggest that use of slope values of F0 may be helpful as additional, if not critical, cues for consonant manner classification. In addition, it is also likely that vowel duration, apart from the duration of the consonant itself, can be a supplementary cue for tense sound identification, though this will not be further investigated in the current study.

5. Conclusion

Two aspects of segmentally affected F0 contours are investigated in this study. First, contrary to some previous studies, the shape of the contour appears to be close to a pure fall regardless of the type of the preceding consonant. So it seems that at least in Korean, there is no rise-fall dichotomy of segmentally influenced F0. This means that the type of consonant cannot be better identified, at least in Korean, even if the segmental F0 is employed as an acoustic cue. It is, however, too premature to discard as a whole the influence of consonant type in determining the shape of F0 contours. Only the three way distinction of Korean — lenis/fortis/aspirated — has been dealt with in the current study and other ways of strong-weak distinction have to be further confirmed.

The second finding of this study is the possibility of using slope as a useful cue to identify consonant types. The overall slope of F0 after aspirated sounds is found to be significantly steeper than that after consonant of other types. It reflects the agreed upon segmental effects of F0: strong consonants like aspirated stops induce a the higher F0 than weaker consonants. Yet further experiments in various other environments are needed to find whether the slope difference is enough to be exploited practically employed as in improving automatic speech recognition or synthesis systems.

Acknowledgement

This study is a modification of a part of Jang (2000b). The overall structure is repeated but major refinement has been implemented with modified experimentation and further discussion.

References

Entropic. 1998. *ESPS/Waves+with EnSig Manual.* Entorpic.

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

Haggard, M., Ambler, S. & Callow, M., 1970. "Pitch as a voicing cue." The *Journal of the Acoustical Society of America*, 47, 613-617.

Han, M. & Weitzman, R. S., 1967. *Studies in the phonology of Asian languages V: Acoustic features in the manner-differentiation of Korean stop consonants.* Technical report, University of Southern California.

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

Hardcastle, W. J., 1973. "Some observations on the tense-lax distinction in initial stops in Korean." *Journal of Phonetics* 1, 263-272.

Hombert, J.-M., 1978. "Consonant types, vowel quality, and tone." in V. Fromkin (Ed.) *Tone: A linguistic survey*, 77-111. Academic Press.

House, A. S. & Fairbanks, G., 1953. "The influence of consonant environment upon the secondary acoustical characteristics of vowels." *The Journal of the Acoustical Society of America*, 25(1), 105-113.

Jang, T. Y., 2000a. "Fundamental frequency in manner differentiation of Korean stops and affricates." *Korean Journal of Speech Sciences*, 7(1), 217-232. The Korean Association of Speech Sciences.

Jang. T. Y., 2000b. *Phonetics of segmental F0 and machine recognition of Korean speech.* Ph.D. Dissertation. University of Edinburgh.

Jun, S.-A., 1996. "Influence of microprosody on macroprosody: a case of phrase initial strengthening." *UCLA Working Papers in Phonetics*, 92, 97-116.

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 stops." *Word*, 21(3), 339-359.

Kim, K.-N., 1968. "F0 variations according to consonantal environments." Ms. University of California at Berkeley.

Kim, M.-R., Beddor, P. S. & Horrocks, J.. 2002. "The contribution of consonantal and vocalic information to the perception of Korean initial stops." *Journal of Phonetics*, 30(1), 77-100.

Kohler, K. J., 1982. "F0 in the production of lenis and fortis plosives." *Phonetica*, 39, 199-218.

Lea, W. A., 1980. "Prosodic aids to speech recognition." in W. A. Lea (Ed.) *Trends in Speech Recognition*, 166-205. Prentice Hall.

Lehiste, I. & Peterson, G. E., 1961. "Some basic considerations in the analysis of intonation." *The Journal of the Acoustical Society of America*, 33(4), 419-425.

Löfqvist, A., 1975. "Intrinsic and extrinsic F0 variations in Swedish." *Phonetica*, 31, 228-247.

Mohr, B., 1971. "Intrinsic variations in the speech signal." *Phonetica*, 23, 65-93.

Ohde, R. N., 1984. "Fundamental frequency as an acoustic correlate of stop consonant voicing." *The Journal of the Acoustical Society of America*, 75(1), 224-320.

Silverman, K. E. A., 1986. "F0 segmental cues depend on intonation: The case of the rise after voiced stops." *Phonetica*, 43, 76-91.

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

Talkin, D., 1995. "A robust algorithm for pitch tracking (RAPT)." in K. K. Paliwal (Ed.) *Speech Coding and Synthesis*. Elsevier.

Taylor, P., 2000. "Analysis and synthesis of intonation using the tilt model." *The Journal of the Acoustical Society of America*, 107(3), 1697-1714.

Taylor, P., & Black, A., 1994. "Synthesizing conversational intonation from a linguistically rich input." in *Proceedings of the 2nd ESCA/IEEE Workshop on Speech Synthesis*.

Terken, J., 1995. "The perceptual relevance of micro-intonation: Enhancing the voicing distinction in synthetic speech by means of consonantal F0 perturbation." in L. Hunyadi, M. Gósy, & G. Olaszy (Eds.) *Studies in Applied Linguistics*, 2, 103-124. Department of General and Applied Linguistics, Lajos Kossuth University of Debrecen, Hungary.

Umeda, N., 1981. "Influence of segmental factors on fundamental frequency in fluent speech." *The Journal of the Acoustical Society of America*, 70, 350-355.

Received : February 10, 2004
Accepted : February 28, 2004

▲ Tae-Yeoub Jang
Department of English
Hankuk University of Foreign Studies
270, Imun-dong, Dongdaemun-gu, Seoul, 130-791, Korea
Tel: +82-2-961-4770 (O), +82-2-535-3637 (H)
Fax: +82-2-965-2183
E-mail: tae@hufs.ac.kr