MODELING QUANTITATIVE VARIATION

- In the Kyungnam Dialect of Korean -

Yong-Hyung Cho[*]

<Received: January, 29, 1997>

ABSTRACT

The objectives of this paper are to see how the declination is realized in the different positions/lengths of the utterance, to see if the F_0 value throughout the utterance changes in a predictable way, and if so, to find out the best quantitative model which fits the declination. The experiment results are as follows. First, the peak value over the utterance can be affected by the position of the peak and length of the utterance. Second, the choice of quantitative models is dependent on the different list lengths. Third, in everyone's speech, there is a baseline (the lowest F_0 value a speaker can use), and the F_0 will not fall below the baseline. Forth, the peak F_0 of the last word in each list shows little variation in pitch value (target F_0) while the number of words in the list affects the starting F_0 values.

1. Introduction

Cho (1994b) shows how interpolation can be used both to reduce the postulated tonal inventory for the Kyungnam Dialect of Korean (KNDK hereafter), and also to make more accurate phonetic claims about the relation between phonological tones and the actual course of F_0 values. In this paper, I pursue the second of these themes by investigating declination phenomena.

Pierrehumbert (1980), and Liberman and Pierrehumbert (1984) reported that in English there is a gradual downdrift and narrowing of the pitch range, which occurs either within the intonational phrase or over the course of utterance: this phenomenon is given the name "declination". This declination has also been studied by other researchers in various languages: Thorsen (1980) for Danish, 't Hart and Cohen (1973) for Dutch, Vaissierre (1971) for French, Fujisaki, et al. (1979), Pierrehumbert and Beckman (1989) for Japanese, Bruce (1977) for Swedish. KNDK is not an exception to this trend, as we will see shortly.

---

[*] Dong-Pusan Technical College

The organization of this paper is as follows. First, the experiment materials and procedures are described, and then quantitative models for the phonetic realization of the declination and of the interaction of declination with pitch prominence in KNDK are introduced.

2. Experiment

This experiment was designed to see how the declination is realized in the different length of the utterance, to see if the F_0 value throughout the utterance changes in a predictable way, and if so, to find out the best quantitative model which fits the declination. Studying the declination pattern would allow us to address possible position dependent effects inside a single intonational phrase or throughout the utterance. We can also look for similarities and differences between English and KNDK since the experimental method is similar to that of Liberman and Pierrehumbert (1984): their experiment involves downstepping - iterative shrinking of pitch range - using contours on five different lists of berry names, and the result of this paper shows that the F_0 of the following berry name in the list is a constant fraction of the F_0 of the preceding berry name.

2. 1. Materials

Five different names of trees were used to produce the declination of F_0 throughout the utterance as shown in Figure 1. The list length varied from two to five and the pitch range remained the same. The tree names were drawn from the set *caŋ.mi.na.mu* 'rose tree', *o.doŋ.na.mu* 'paulownia tree', *eŋ.du.na.mu* 'cherry tree', *sa.gwa.na.mu* 'apple tree', and *p^{h}o.do.na.mu 'grape tree'. Order of the words was permutated to have all words of equal* importance in the list and to minimize possible semantic effects, and consonantal influences on F_0.

2. 2. Procedures

A woman and a man participated in this experiment. Both were native speakers of KNDK. Each subject was given a stack of note cards containing a list of two to five tree names. The desired intonation pattern was demonstrated to subjects and they were instructed to begin uttering each list with roughly the same pitch. If subjects had only been asked to say the words in the list without a demonstration of the desired intonation pattern, the pitch contour of each word could have been almost the same throughout the utterance. The other sources of variation in each experiment were restricted in order to keep the task manageable, and the ability of subjects to produce the desired intonation pattern naturally was examined before the beginning of actual recordings.

Audiotaped recordings were made in a sound treated room and audiotaped words were recorded into the computer. The recorded utterances were digitized and F_0 contours were extracted using Signalyze^{TM}.

Recall that each tree name has the form $\sigma \sigma +$ namu. Since the syllable na in namu is the syllable associated with the word's H tone, the peak F_0 value was measured and associated with the word as a whole. The fixed position of the most prominent syllable is very helpful in examining the positional variation of F_0 values because the position of the highest F_0 in each word remains the same. If the position of the F_0 peak in each word varied, it would be hard to study the positional variation of F_0 in the list: suppose that the F_0 peak of the third word in the five-word list is in the fourth syllable of the word, and the F_0 peak of the third word in a different five-word list is in the first syllable of the word; then the measured F_0 of the former would be lower than that of the latter because the impact of declination would be greater on the fourth syllable of the word than the first syllable of the word at the same position of the list. The F_0 values in each tree name are not affected much by consonants because word-internal consonants are voiced.

2. 3. Results

Figure 2 and Figure 3 show the results of the experiment from the two subjects. The x-axis in these figures shows the peak position in the list from the first, and the y-axis shows the F_0 in Hz. The lines running through the plotting points in each figure is called a trend line, or a regression line. Each regression line in these figures indicates a different length of the list, and the plotting points on each line shows the average F_0 values of the peak of tree names at different positions in the list. The plotting point △ is used for the two-word list, ☐ for the three-word list, ◇ for the four-word list, and x for five-word list.

As all the regression lines in these figures show, the F_0 decreases as the position of the peak nears the end of the list. The F_0 of the second peak is lower than the F_0 of the first peak, and the F_0 of the third peak is lower than the F_0 of the second peak, and so on. If we take the five-word list as the subject MRR for example, the averaged F_0 value of each peak, counting from the first peak in the list, is 276 Hz, 243 Hz, 220 Hz, 210 Hz, and 189

Hz, in sequence. The F_0 of each peak decreases as the position of the peak moves to the end of the list. This is well illustrated in Table 1, in which mean F_0 (in Hertz) is the averaged value of 20 measurements for each list.

At this point, we can say that the F_0 value of each peak is position dependent. That is, as a peak is closer to the first peak of the list, the F_0 of that peak is higher than the F_0 of following peaks, if any. However, it should be remembered that the F_0 of the second position, for example, in the list would vary depending on the length of the list. That is, the F_0 of the second peak in the three-word list would not be the same as the F_0 of the second word in the five-word list. The F_0 of the second peak in the two-word list would not be higher than the F_0 of the third peak in the five-word list, either. For example, the F_0 of the second peak in the three-word list is 25 Hz lower than the F_0 of the second peak in the five-word list, as shown in Table 2 for subject MRR. The biggest difference occurs when we compare the second peak of the five-word list and that of the two-word list.

The situation would not be very different, even if we count the position of the peak from the end of the list. The F_0 of the penultimate peak in the five-word list would not be the same as the F_0 of the penultimate peak in the three-word list. In most cases, the F_0 of the penultimate peak in the three-word list would be higher than the F_0 of the penultimate peak in the five-word list, as shown in Table 3. If we take the subject MRR example, the F_0 of the penultimate word in the three-word list is 8 Hz higher than the F_0 of the penultimate word in the five-word list.

Now, it becomes clear that the F_0 value of each peak is not only position dependent in the list but also length dependent. This view is also supported by the F_0 of the first peak in each list. As the number of the word in the list increases, the starting F_0 value, which is the F_0 of the first word in the list, also increases accordingly. This tendency is also observed in English (Liberman and Pierrehumbert 1984). In addition to these interesting observations, there is another point that draws our attention. The peak F_0 of the last word in each list shows little variation in pitch value. The number of words in the list and the starting pitch value of the first word in the list have no effect. To accurately define the degree of variation in pitch, we can use the standard deviation which indicates the typical amount by which values in the data set differ from the mean. The standard deviations shown in Table 4 were obtained from all of the starting F_0 values and all of the last F_0 values in all different list lengths. As we predicted, the standard deviation of the starting F_0 of each list is much greater than that of the last F_0 peak - the standard deviation of all of the last peaks in each subject is less than 2 and greater than 1, whereas the standard deviation of the all of the first peaks is 17.43 Hz and 6.22 Hz for the subject MRR and

YJC, respectively. The fact that the F_0 of the last peak does not vary much without regard to different list length contrasts with Liberman and Pierrehumbert's (1984) report on English, where the peak F_0 of the last word in shorter list is higher than that of the last word in longer list.

3. Quantitative Model

As shown in Figure 2 and Figure 3, each line connecting plotting points forms different falling contours. The line for the two-word list is a straight line and the line for the five-word list is a curved line concave upward. As the number of words (length) increases in the list, the trendline moves closer to a curved line from a straight line. As the intervening numbers of the word between the first peak and the last peak in the list increase, the slope of the trendline becomes less steep and the difference in F_0 between adjacent peaks becomes smaller. For example, the slope of the three-word list is steeper than that of five-word list, and the F_0 difference between adjacent peaks in the three-word list should be larger than the F_0 difference between adjacent peaks in the five-word list. The fact that the contours of each trendline and the slopes of each trendline are different among different list lengths leads us to believe that different quantitative models for each trend line are needed in order to explain the different trends of declination. Then, it will be very intriguing to find whether the F_0 value in each regression line decreases in a predictable way, what type of a model best fits each trendline, and the degree to which a selected model can explain the downward trend in F_0 throughout the utterance. The following section will seek the answers to these questions.

3.1 Possible Models for Each Trendline

As illustrated in section 0, there is a relationship between the peak number and the F_0 of the peak: the F_0 values on the y-axis in Figure 2 and Figure 3 decrease as the peak number on the x-axis increases. Regression analysis is suitable for the task to find out the relationship between the peak number and the F_0 value of the peak since regression analysis shows the relationship between one variable and another set of variables.

If we assume that a regression trendline is a straight line, there should be a linear relationship between x and y. If the trend is curved, there is a curvilinear relationship between x and y. When there is a linear relationship between x and y, the regression line can be represented by a linear equation of the form:

(1) $y_i = \beta_o + \beta_1 x_1$

where y_i is the response variable, x_1 is a regressor variable, and β_o and β_1 are unknown parameters to be estimated. The two constants β_o and β_1 in the equation determine the location and the slope of the line. The constant β_o is the intercept, that is, the value of y when the line crosses the vertical axis (y-axis). The constant b1 is the slope of the line, that is, the change in y that corresponds to a one-unit increase in x. We will name the formula in (1) a straight line model.

When there is a curved relationship between x and y, either one of the following equations can be used:

(2) $y = \beta e^{-xb}$

where x is a variable, β and b are constants, and $2<e<3$, or

(3) $y = \beta x^{p}$

where x is a variable, $\beta$and ^{p} are constants, and $r < 0$.

Unlike a linear regression model, a curved model is complicated and there might be other possible models. However, two possible equations - the exponential equation in (2) and the power equation in (3) - are explored in the present study. We will name the equation in (2) an exponential model, and the equation in (3) a power model. Now that there are three models to compare, the next task is to apply these models to the actual data and to find a model which fits the data best.

Let us first start with a three-word list. Figure 4 shows three prediction lines for a three-word list of the subject MRR. A solid line crossing plotting points ◇ represents a straight line produced by the straight line model, the dotted line crossing plotting points ☐ represents an exponential curve produced by exponential model, the dotted line crossing plotting points x represents a curve produced by the power model, and unconnected ● plotting points represent actual data points for each peak. When we look at the three lines produced by three different models, it can be seen that power model is not a suitable prediction model since the second predicted point is the farthest from the actual data point. However, it is very hard to determine which model best fits the actual data points "by eye" between the straight line model and the exponential model. To resolve this problem, we need to examine the R2 value, which is an indicator of how much variation in the data is explained by the model. R squared values in Table 5 shows that the exponential model and the straight line model in each subject are almost perfect models for the three-word list. For subject MRR, the straight-line model is marginally (0.01) better than the exponential model for subject MRR. For subject YJC, the exponential model predicts the data 0.003

better than the straight line model. For each subject, it is appreciated to use a straight line model for the best prediction model for three-word list for each subject because the difference between the two models is not statistically significant, and there is a linear relationship between the F_0 values of peaks and peak numbers.

Figure 5 shows three prediction lines for the four-word list of the subject MRR. A dotted line crossing plotting points ◇ represents a straight line produced by the straight line model, a dotted line crossing plotting points □ represents an exponential curve produced by exponential model, a solid line crossing plotting points x represents a curve produced by the power model, and unconnected plotting points ● represent actual data points for each peak. As can be seen in this figure, the power model predicts the data best and its R squared value in Table 6 is the highest among the models. However, subject YJC shows quite a different result. The power model has the lowest R squared values and is therefore the most unsuitable model. Instead, the straight line is the best best predictor for the four-word list. Alhough it is possible to use the exponential model for the four-word list of this subject, we will adopt the straight line model for the reason just mentioned.

So far, each prediction model for three-word list and four-word list has been tested, and it has been found that the straight line model is the most suitable model for each of the different list lengths except in the case of the four-word list for the subject MRR. Now, we will examine if the straight line model is still the most suitable model for the five-word list.

Figure 6 shows three prediction lines for the five-word list of the subject MRR. Plotting points are used in the same manner as in other figures; ◇ for the straight line model, □ for the exponential model, and x for power model. It is difficult to determine by eye which model best fits the actual data points. If the fifth peak had not been considered, the power model could have been the best model for the five-word list. However, due to the final fall of the fifth point (The power model predicts the fifth peak 6 Hz is higher than the actual fifth peak (189 Hz), the exponential model predicts that the five-word list is best for both subjects as shown in Table 7. If this is the case, it would be easy to obtain each peak value in the list once we get a downstep constant (a constant which determines the successor relation among peaks in a series). Since the peaks following the first peak in the exponential model is a constant fraction of its precusor, we can obtain a downstep constant by dividing the following peak value by preceding peak value as shown below:

(4) $c = X_{i+1}/X_i$

where c is a downstep constant less than 1.

The downstep constants, obtained from equation (4), for the subject MRR and the subject YJC are 0.91 and 0.90 respectively, in the five-word list. In order to obtain the F_0 value of the second peak, for example, we simply multiply the downstep constant by the first peak.

So far, list lengths except two-word lists have been examined. The two-word list was excluded from examination as it could not indicate whether it fits more suitably to the straight line or to the curved line . That is, if there is only two data points in the graph, whether a straight line or a curved line can cross the second data point, giving us 100% predictability by either model.

4. Estimating the Fifth Peak

In the previous section, the best quantitative model for each list was examined. In this section, It will be examined how accurately a model obtained from F_0 values of the first four peaks or the first three peaks in the five-word list can estimate the F_0 value of the fifth peak.

Let us first estimate the fifth peak values using the first four peaks in the five-word list. To do this task, we need to obtain equations for each model using the first four peak values, and then, it is possible to estimate the fifth peak values from the equations. The plotting point x in Figure 7 is the F_0 of the fifth peak obtained from the power model, and is 6 Hz higher than the actual F_0 value of the fifth peak. The plotting points ◇ and □ are 187 Hz and 189 Hz, respectively, and the former obtained from the straight line model and the latter from the exponential model. The fifth peak value predicted from the exponential model is the same as the actual F_0 of the fifth, 189 Hz. Here again, the exponential model turned out to be the best model for the five-word list.

Since it is proved that the exponential model which best fits the five-word list is also the most suitable best to predict the fifth peak value using the first four peaks, let us examine which quantitative model best estimates the fifth peak in the five-word list for the first three peak values.

Plotting points ◇, □, and x in Figure 8 show the predicted values for each model. There is more than a 10 Hz difference between the actual pitch value and the predicted values. This result indicates that it is not plausible to use the first three peak values to predict the fifth peak values in the five-word list.

Table 8 is a summary of the predicted F_0 for the subject YJC. In this table, the F_0 (115 Hz) predicted from the exponential model from the first four peaks is the closest to the actual F_0 value (116 Hz). It is also shown that the predicted F_0 using the first three peak values significantly deviates from the actual peak value.

In this section, it is shown that quantitative models obtained from the first three peak values in the five-word list does not actually predict the 5th peak. This is because the trendline moves upward from the fourth peak in the five-word list, while the first three peaks fall in a linear regression fashion.

5. Conclusion

So far, it has been shown that the peak value over the utterance can be affected by the position of the peak and length of the utterance, and that the most suitable quantitative model is dependent on the different list lengths. If we do not find the best quantitative model, it is difficult to determine the peak value at a particular position of the utterance. To obtain the best quantitative model, peak values at each position in the utterance and a downstep constant are needed. Once, these values are determined the best model and the extent to which actual peak values deviate from the predicted values by the model can be seen. In this framework, the realization of the peak values did not depend on the reference line. In Liberman and Pierrehumbert's study (1984) of English, the F_0 values for peaks are specified in terms of distance above a reference line. The distances of H tones above the reference line can vary due to declination. The effect of declination is as follows:

(5) $X_{i+1} = r + c(X_i - r)$

where c is the downstep constant, less than 1, and r is reference line.

Liberman and Pierrehumbert suggest that knowledge of any one of the peak F_0 measurements in a given utterance allows us to predict all other peak F_0 measurements in that utterance. However, there are two questions arising from this statement. First, how can a downstep constant be determined without knowing which model will best fit the data? The downstep constant they used was based on the assumption that the declination is an exponential decay. As shown in the preceding sections, it is impossible to tell which model would best fit until all possible models have been tested. Depending on the model, the downstep constant will be different. Second, the high-tone value and the reference line can vary independently as shown in Pierrehumbert and Beckman (1989), making it very difficult to determine a reference value r. When attempting to predict the peak values using equation (5), a gross error as shown in Table 9. In this instance, predicted F_0 values are much higher than actual values. Accordingly, the fifth data point in the five-word list falls below the value predicted by their model in Liberman and Pierrehumbert's study (1984) of English. However, the values predicted from our model show only a relatively small amount of error, as displayed in Table 10.

References

Bruce, G. 1977. *Swedish Word Accents In Sentence Perspective.* Travaux de l'Institute de Linguistique de Lunc, CWK Gleerup, Malmo.

Cho, Y. H. 1994. "The Interaction of Lexical Tone and Non-Lexical Tone in the KyungNam Dialect of Korean." Proceedings of the Eleventh Eastern States Conference on Linguistics. Ithaca, NY: DMLL Publications.

Fujisaki, H., Hirose, and K. Ohta. 1979. "Acoustic Features of the Fundamental Frequency Contours of Declarative Sentences in Japanese." Annual Bulletin No. 3, Research Institute of Logogpedics and Phoniatrics, University of Tokyo.

Liberman, M. and J. Pierrehumbert. 1984. "Intonational Invariance under Changes in PitchRange and Length." In M. Aronoff and R. T. Oehrle, eds., Language and Sound Structure: Studies in Phonology presented to Morris Halle, 157-233. MIT Press, Cambridge, Massachusetts.

Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation.* Doctoral Dissertation, MIT, Cambridge, Massachusetts.

Pierrehumbert, J. and M. Beckman. 1989. *Japanese Tone Structure.* MIT Press, Cambridge, Massachusetts.

'tHart J. and A. Cohen. 1973. "Intonation by Rule: a perceptual quest," *Journal of Phonetics 1,* 309-327.

Thorsen, N. 1980a. "Intonation Contours and Stress Group Patterns in Declarative Sentences of Varying Length in ASC Danish." Annual Report of the Institute of Phonetics, University of Copenhagen 14, 1-29.

Thorsen, N. 1980b. "A Study of the Perception of Sentence Intonation - Evidence from Danish." *Journal of Acoustical Society of America 67,* 1014-1030.

Vaissierer, J. 1971. *Contibution a la synthese par regles du francais,* Doctoral dissertation, Universite des Langues et Lettres des Grenoble.

▲ 640 Pansong-Dong, Heaundae-Gu, Pusan, 612-715, Korea
Dong-Pusan Technical College 612-715
Tel : (051) 520-3892(O) FAX: (051) 520-3844
(051) 782-9030(H)
e-mail : yhcho@sulbong.cesb.ac.kr

APPENDIX

Table 1. Mean F_0 (Hz) for each peak

Subject MRR	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5
Two-word list	229	190			
Three-word list	241	218	187		
Four-word list	258	222	208	190	
Five-word list	276	243	220	210	189
Subject YJC					
Two-word list	157	117			
Three-word list	164	136	116		
Four-word list	167	153	136	119	
Five-word list	172	154	138	6128	116

Table 2. Differences (in Hz) between means of different list lengths, by position, counting from the first to the last. The numbers in the first column represent the lengths of the two lists being compared.

Subject MRR	Position 1	Position 2	Position 3	Position 4
5-4	18	21	12	20
5-3	35	25	33	
8 5-2	47	53		
Subject YJC				
5-4	5	1	2	9
5-3	8	18	22	
5-2	15	37		

Table 3. Differences (in Hz) between means of different list lengths, by position, counting from the last to the first. The numbers in the first column represent the lengths of the two lists being compared.

Subject MRR	Position 1	Position 2	Position 3	Position 4
5-4	-1	2	-2	-15
5-3	2	-8	-21	
8 5-2	-1	-19		
Subject YJC				
5-4	-3	-8	-15	-13
5-3	0	-8	-26	
5-2	-1	-29		

Table 4. Standard Deviation for the first peaks and the last peaks.

	The First Peak	The Last Peak
Subject MRR	17.43	1.59
Subject YJC	6.22	1.27

Table 5. *R squared* values and equations of each prediction model for three-word list

	Model	**Equation**	**R^2**
Subject MRR	Straight	$y = -27.274x + 269.9$	.99
	Exponential	$y = 276.64e^{-0.128x}$	.98
	Power	$y = 244.81x^{-0.2241}$	.93
Subject YJC	Straight	$y = -24.079x + 186.64$	.99
	Exponential	$y = 194.12e^{-0.174x}$	.99
	Power	$y = 165.14x^{-0.3118}$	.98

Table 6. *R squared* values and equations of each prediction model for four-word list

	Model	**Equation**	$\mathbf{R^2}$
Subject MRR	Straight	$y = 21.789x+274.14$	.95
	Exponential	$y = 279.05e^{-0.0982x}$	.97
	Power	$y = 258.54x^{-0.213}$	.99
Subject YJC	Straight	$y = -16.31x + 184.37$	.998
	Exponential	$y = 189.86e^{-0.115x}$	.99
	Power	$y = 172.04x^{-0.2379}$	.92

Table 7. *R squared* values and equations of each prediction model for five-word list

	Model	**Equation**	$\mathbf{R^2}$
Subject MRR	Straight	$y = -20.62x+289.67$	.96
	Exponential	$y = 295.91e^{-0.0899x}$	.98
	Power	$y = 279.73x^{-0.2231}$	.97
Subject YJC	Straight	$y = -13.675x+182.58$	.98
	Exponential	$y = 187.24e^{-0.1964x}$	.99
	Power	$y = 175.89x^{-0.2367}$	.96

Table 8. The F_0 of the fifth peak predicted from the first four peaks and the first three peaks in the five-word list.

Subject YJC (116 Hz)	Fo predicted from the first four peaks	Fo predicted from the first three peaks
Straight Model	111	103
Exponential Model	115	110
Power Model	127	126

Table 9. Errors from equation (5): Predicted F_0 - Actual F_0

	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5
Subject MRR		24	18	7	19
Subject YJC		13	12	7	11

Table 10. Errors from equation (5): Predicted F_0 - Actual F_0

	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5
Subject MRR	-2	1	2	-1	0
Subject YJC	-5	4	5	-4	-1

Figure 1. An Fo contour for the tree list caŋminamu 'rose tree', odoŋnamu 'paulownia tree', eŋdunamu 'cherry tree', sagwanamu 'apple tree', phodonamu 'graph tree', showing the declination of Fo throughout the utterance.

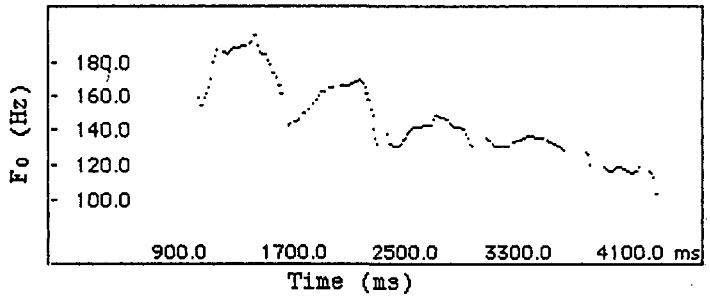

Figure 2. Declination Data for Subject MRR. (X-axis = peak number, Y-axis = Fo in Hz)

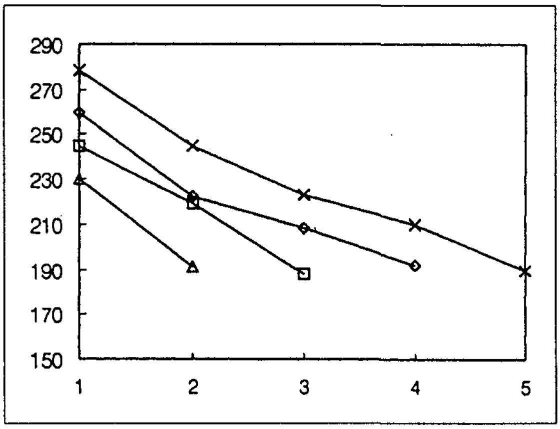

Figure 3. Declination Data for Subject YJC

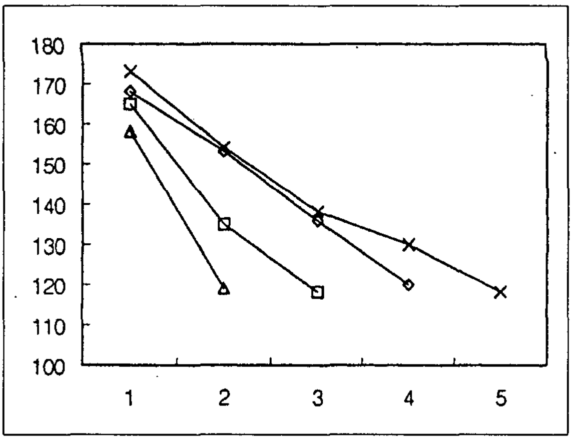

Figure 4. Each line is the prediction line produced by three different models for the three-word list for the subject MRR, and black circles are actual pitch values.

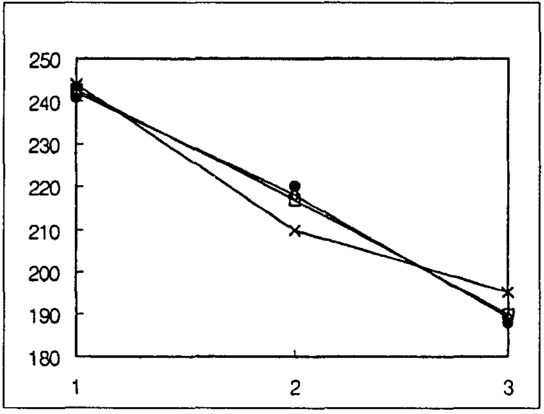

Figure 5. Each line is the prediction line produced by three different models for the three-word list for the subject YJC, and black circles are actual pitch values.

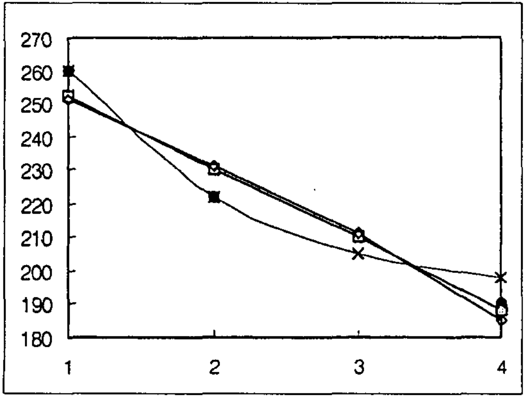

Figure 6. Each line is the prediction line produced by three different models for the five-word list for the subject MRR, and black circles are actual pitch values.

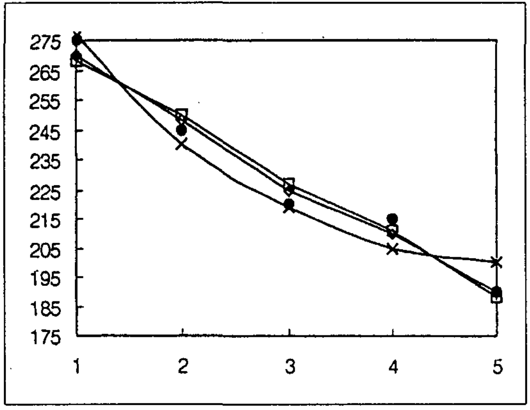

Figure 7. Plotting points. △ and X are predicted values for each model using the first four peak values in the five-word list. Black dots are actual peak values.

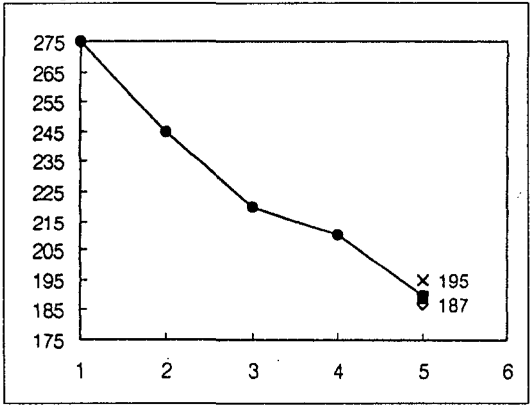

Figure 8. Plotting points. △ and X are predicted values for each model using the first three peak values in the five-word list. Black dots are actual peak values.