Vowel intrinsic pitch in Standard Chinese

Shi Bo and Zhang Jialu

Institute of acoustics, Academia Sinica

ABSTRACT

We investigated whether an intrinsic pitch (IP) effect occurs in Standard Chinese (SC) and if it exists how IP and pitch level interact with each other. The fundamental frequencies (F0) of each 9 Chinese vowels at different tonal points were measured in three positions: (1) in a monosyllable, (2) in the word-initial and (3) the word-final position of a disyllabic word. The test items (400 monosyllables and 509 disyllabic words) were embedded in a frame sentence and uttered by 5 male and 5 female informants. The results show that the characteristics of IP are to be found in all different tones of SC in spite of the fact that those tones have different tonal configurations. Further, the higher the relative pitch value, the larger the difference in F0 among the vowels. The IP differences are reduced in word-final position. These results suggest a new hypothesis.

INTRODUCTION

Intrinsic pitch (intrinsic F0) describes the influence of tongue height of vowels on the F0-value associated with them: high vowels have higher average F0-values than low vowels when other factors are kept constant.

A great deal of research has been devoted to the analysis and quantification of intrinsic pitch in several languages: English (Crandall,1925; Taylor, 1933; House and Fairbanks,1953;

Lehiste and Peterson, 1961), Italian (Ferrero et al., 1975), Danish (Petersen, 1978), Japanese (Nishinuma, 1977), French (Di Cristo, 1985), German (Neweklowsky, 1975), Greece (Samaras, 1972), Taiwanese Chinese (Zee, 1978), Yoruba (Hombert, 1977), Serbo-Croatian (Ivić and Lehiste, 1963), Itsekiri (Ladefoged, 1964), and Chinese (Connel et al., 1983). IP has also been observed when vowels were sung at the same pitch (Ewan, 1979. Personal communication by C.K.Chang).

Various experimental conditions were applied in these studies. In the early experiments, isolated "real" words (such as in Peterson and Barney, 1952) as well as "nonsense" words (such as in House and Fairbanks, 1953) were used. The segmental environments (i.e. consonantal context) were carefully controlled. Later, the test words were embedded in a frame sentence (as in Lehiste and Peterson, 1961: "Say the word ___ again."). The effects of prosodic environment on IP were also studied. Petersen (1976) reported that the magnitude of IP in stressed syllables is larger than in unstressed syllables. Similar results were obtained for Italian accent/nonaccent words (Ferrero and al., 1975). All of these studies generally showed similar results except Umeda's (1981) which reported that there were no consistent IP effects in a 20-min reading by two speakers. In order to investigate whether IP effects occur in connected speach, Ladd and Silverman (1984) compared test vowels (in German) in comparable segmental and prosodic environments under two different experimental conditions: (1) a typical laboratory task in which a carrier sentence served as a frame for test vowels; (2) a paragraph reading task in which test vowels occurred in a variety of prosodic environments. It was shown that the IP effect does occur in connected speech, but that the size of the IP differences is somewhat smaller than in carrier sentences. They pointed out that Umeda's finding was questionable because she apparently had not made any attempt to control for the prosodic environment of the vowels that were measured. In a recent study, Shadle (1985) investigated the interaction of IP and intonation in running speech. She examined the F0 of the vowels [i,a,u] in four sentence positions. The results showed a large main effect of IP that lessened in sentence final position. The study by Zee (1978) on Taiwanese Chinese showed that the IP also appears in a tone language, and that its magnitude is less for lower tones. In his study only two contrasting tones, high tone and low tone were analyzed.

However, none of these studies were concerned with the roles of pitch level and the position in the word in affecting intrinsic pitch. The main goal of the present experiment was to get a general idea about the effect of intrinsic pitch in Standard Chinese. The effect was to be studied as a function of the following variables: (1) pitch level (in different tones); (2) position in disyllabic words (word-initial and word-final).

METHOD

The material consists of two parts, 400 monosyllables and 509 disyllabic words. The monosyllables consist of all possible arrangements of consonants and vowels in Chinese, each arrangement having all four tones. Among them there are 279 "real" monosyllabic words and 121 "nonsense" words. The disyllabic words consist of one test syllable (a vowel preceded by an initial consonant) and one matched syllable. The matched syllable was chosen in such a way that the test vowels could be compared in a similar segmental environment and the same tonal surroundings. Examples are fāhuà/fūhuà; wēibā/wēibō/wēibī; tújǐng/tíxǐng, (The test syllables are underlined). Of the test syllables 273 were in word-initial and 236 in word-final position. As many combinations of two tones as possible were involved in this part.

In order to make all test items occur in the same phonetic environment and approach the situation of connected speech, all the monosyllables and disyllabic words were embedded in the frame sentence / Wǒ dú _ zì./ "I utter the character _." and / Wǒ dú _ _ zhè ge̊ cí./ "I utter the word _ _." respectively.

10 speakers (5 males and 5 females) of Standard Chinese were recorded. Before the recordings each speaker was given a short training period . A natural style was aimed at. The test materials were read once by each speaker in an acoustically treated room.

The records were fed into a Visi-Pitch (model 6087) for the extraction of F0. The counter provides a digital display of F0 for sustained vowels while the cursor allows the user to determine the F0 of any point on the pitch curve shown on the screen with +1 Hz accuracy.

Fig.1 shows the measuring points of F0. They are: for high tone (T1) the middle point T1; for rising tone (T2) the lowest point T2-1 and the highest point T2-2; for dipping tone (T3) the starting point T3-1 and the lowest point T3-2; for falling tone (T4) the highest point T4-1 and the lowest point T4-2.

As a first step in our analysis we only cared about average IP differences between vowels but ignored the differences between consonantal context and interspeaker variabilities. The statistical method was a one-way analysis of variance (with speakers and consonantal environments as a repeated measure).

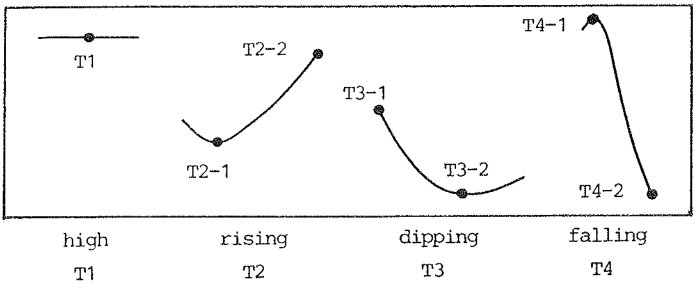

Fig.1 Measuring points of fundamental frequency

RESULTS

1. Vowel intrinsic pitch in four tones

The data which will be analyzed in this section were derived from 400 monosyllables. The intrinsic F0-values for each of 9 vowels at different tonal points are given in Table 1 in which the data are mean values averaged across consonants, for 5 males and 5 females respectively. The vowels were arranged in order of F0-value (based on the data in Table 1 averaged across 10 speakers) in Table 2. The relative mean F0 differences between low vowel [a] and the remaining vowels are shown in Table 3 (see `monosy.` parts).

The data mentioned above permit us to make the following observations: 1). At points T1, T2-2, and T4-1 the F0-values of high, middle and low vowels go from high to low and the F0 differences between high and low vowels are 10-30 Hz; 2). At points T2-1 and T3-2 a high vowel also has a higher F0 except that the F0-value of [o] is a bit higher than that of [ʅ] and [i]. This is also shown graphically in Fig.2 (see ·—·) The data at these five points show that Chinese also exhibits the influence of intrinsic pitch.

At points T3-1 and T4-2 the situation is more complex. In the measurements we found considerable inter and intra-speaker variability for F0-values at T3-1. For point T4-2, the main problem is that at the end of T4 the energy is very low and the

Table 1. Mean intrinsic F0-value for each of the 9 Chinese vowels at different tonal points, derived from 400 monosyllables, averaged across consonantal contexts, and for 5 males and 5 females respectively

sex	vowel (IPA)	number of occurrences (each tone)	F0-value (Hz)						
			T1	T2-1	T2-2	T3-1	T3-2	T4-1	T4-2
male	i	12	175	118	167	113	89	197	97
	ɿ	3	181	122	171	116	90	208	99
	ʅ	4	179	116	169	115	90	195	101
	y	6	180	119	175	115	90	197	101
	u	19	181	117	168	112	90	206	105
	e	19	164	114	156	114	88	187	101
	o	18	168	117	160	116	90	184	100
	ɚ	1	170	116	170	122	88	178	100
	a	18	154	111	151	108	83	175	97
female	i	12	291	205	265	219	169	312	180
	ɿ	3	302	206	271	214	172	326	182
	ʅ	4	295	200	264	216	168	319	192
	y	6	300	209	278	219	171	318	176
	u	19	307	209	289	218	172	335	184
	e	19	289	202	270	215	170	315	183
	o	18	278	200	270	213	170	310	183
	ɚ	1	302	200	274	209	161	314	182
	a	18	276	198	255	227	171	302	187

Table 2. Vowels ordered from high F0 to low F0 (derived from 400 monosyllables, averaged across 5 male and 5 female speakers and consonantal contexts)

FO	T1	T2-1	T2-2	T3-1	T3-2	T4-1	T4-2
	u	u	u	y	u	u	ʅ
high	ɿ	y	ʅ	ɿ	ɿ	a	u
	y	ɿ	y	u	y	i	e
	ʅ	ʅ	ɿ	i	o	ʅ	a
↓	i	i	i	o	ʅ	ɿ	o
	e	o	e	ʅ	i	u	ɿ
low	o	e	o	e	e	o	y
	a	a	a	a	a	e	i

Table 3. Relative F0 differences (Δ F0) between the vowel [a] and the remaining 7 vowels at different tonal points in three conditions: monosyllable, word-initial and word-final, (data were averaged across consonantal contexts and all speakers)

		F0 difference (Hz)							
point	position	i	ɿ	ʅ	y	u	e	o	mean 2**
	monosy.	16	28	18	19	32	12	8	19
T4-1	word-i.	16	5	11	15	22	19	18	15
	word-f.	8	10	9	6	16	13	2	9
	mean 1*	13	14	12	13	23	16	9	14
	monosy.	18	27	22	25	30	12	8	20
T1	word-i.	20	24	30	23	21	21	20	23
	word-f.	11	23	8	19	12	15	13	14
	mean 1	17	25	20	23	21	16	14	19
	monosy.	11	16	11	21	23	8	10	14
T2-2	word-i.	22	21	11	21	17	9	7	15
	word-f.	13	-	2	15	14	4	3	10
	mean 1	15	18	8	19	18	7	7	13
	monosy.	7	9	4	9	8	4	4	6
T2-1	word-i.	12	21	12	7	10	3	2	9
	word-f.	11	-	8	5	11	0	2	6
	mean 1	10	15	8	7	10	2	2	7
	monosy.	2	4	2	4	4	2	3	3
T3-2	word-i.	5	7	6	6	5	3	-2	4
	word-f.	2	5	3	1	3	-1	1	2
	mean 1	3	5	4	4	4	1	1	3

* mean for 3 positions
** mean for all 7 vowels

A negative value indicates that the F0-value of vowel was lower than that of vowel [a]

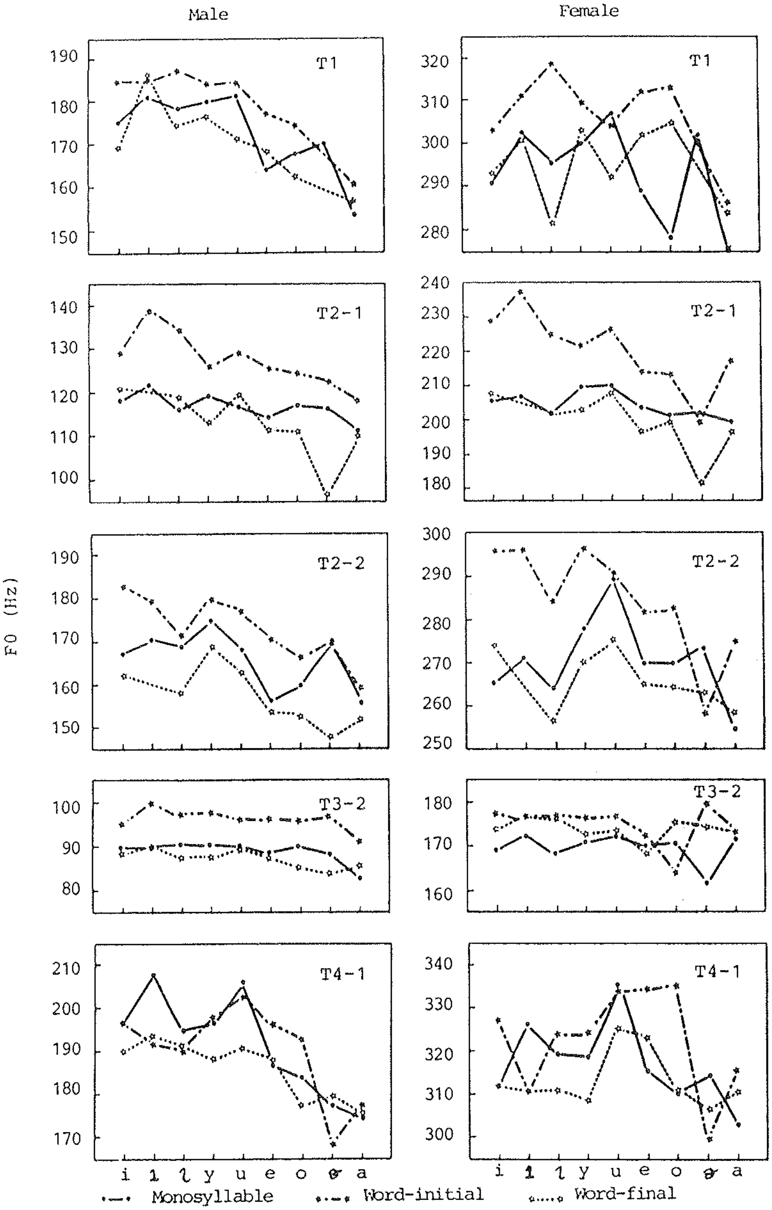

Fig.2 Mean F0 for the vowels, plotted as a function of tongue height, averaged across consonants and for 5 males and 5 females respectively

periodicity is not good enough to permit precision in measurements. As a result there is no consistent influence of IP at these two points.

2. Effect of word-position on IP

When the test syllables were in the disyllabic words there are additional factors influencing F0. There is some modification in F0 caused by the adjacent tone. The data in Fig.2 (★ and☆) and table 3 (see `word-i.` and `word-f.` parts) show that, though meaning and tonal environment are not separated in the data, the effect of intrinsic pitch still occurs regardless of whether the test vowels were in word-initial or word-final position. However the magnitude of IP was reduced in word-final position. This reduction appears to be related to a lowering of F0 in this position (in Fig.2, the curves derived from the word-final position are the lowest ones in most cases).

3. Interaction of intrinsic pitch with pitch level

Fig.2 shows F0-values of 9 simple vowels as a function of the tongue height associated with them. In each part of Fig.2, from left to right, the tongue height of the vowel goes from high to low and it is accompanied by a drop in F0, which reflects the effect of IP. But the curves at different tonal points have different slopes. It indicates that although each vowel appears to be associated with an intrinsic F0 at a certain tonal point, the differences of intrinsic F0 across the vowels vary from point to point. This variance can be seen more clearly in Table 3, in which from top to bottom, the mean relative F0 differences (ΔF0) between the low vowel [a] and the remaining vowels become smaller and smaller as F0 drops. The ΔF0 at points T1 and T4-1 (high F0) are much larger than those of point T3-2 (lower F0).

Going a step further, there is little difference in ΔF0 between the males and the females in spite of the fact that the F0 of the females is higher than that of the males. It indicates that the magnitude of ΔF0 is directly proportional to some kind of relative pitch value rather than to the absolute F0-value. In tone languages, "tonal value" and "tonal register" are often used to discribe the relative relationships of pitch values. If we call the absolute F0-minimum as F0(min) and F0-maximum as F0(max), then the tonal value T(p) (in Oct.) for F0(p) (in Hz) is the binary logarithm of the quotient of F0(p) and F0(min). When F0(p) is equal to F0(max), the T(max) is the tonal register.

Thus

Tonal value: $T(p)=\log_2(F0(p)/F0(min))$ Oct.

Tonal register: $T(max)=\log_2(F0(max)/F0(min))$ Oct.

The ΔF0 between i-a and between u-a are plotted as a function of the normalized tonal value (T(p) divided by T(max)) in

Fig.3. The "●" represents the averages over i-a and u-a, across the males and the females. It is obvious that the higher the tonal value, the larger the ΔF0. In other words, the IP is more marked in the high frequency region of the tonal register than in the lower one. But the slopes of the two curves of the females turn negative when the normalized tonal value is bigger than 0.8. It seems that when the F0-value goes beyond certain limits, the direct proportional relation between ΔF0 and F0 will no longer be tenable. This suggests that it might be worth while to study intrinsic pitch in a larger F0 dynamic range such as in singing.

Table 4. Comparison of FD* for different languages

FD of		FD of Chinese	
		male	female
Italian~	23.3	25.2 (T1)	23.5 (T1)
		24.5 (T4-1)	22.6 (T2-2)
Danish	16.5	14.5 (T2-2)	19.6 (T4-1)
Greece	11.9		
English	10.5		
French	9.8		
German	9.6		
Italian**	9.2	7.2 (T2-1)	9.1 (T2-1)
Japanese	9.0	7.2 (T3-2)	

* FD=FH-FL FH: mean F0-value of high vowels [i,u,y]
FL: mean Fo-value of low vowels [ɛ,œ,ɔ,a]

~ accent
** unaccent

4. Comparison

For the purpose of comparison, the term "FD" was defined as the difference between the mean F0-value of high vowels [i,u,y] (FH) and the mean F0-value of low vowels [ɛ,æ,ɔ,a]) (FL), i.e. FD=FH-FL. The FD-values of some other languages (quoted from Di Cristo, were compared to the FD-values of SC in Table 4. It seems that the magnitude of IP of SC at different tonal points just varies between those of other languages. Italian and Chinese are similar with respect to the fact that for Italian the FD-value of an accented syllable is larger than the FD-value of an unaccented one, and for Chinese the FD-value for a higher tone is larger than the FD-value for a lower tone.

The mean ratios of [i] and [u] to [a] as established in the present study and in a number of other investigations are given in Fig.4. It is clear that the data of the present study are in good agreement with the results of other studies.

DISCUSSION

There have been various hypotheses for the cause of IP. Taylor's "dynamogenetic irradiation hypothesis" (1933) appears to be the earliest attempt. He thought that "the most plausible explanation of the vowel pitch triangle appears to be that it results from dynamogenetic irradiation of the tongue tension to the vocal cord musculature." But it was obviously

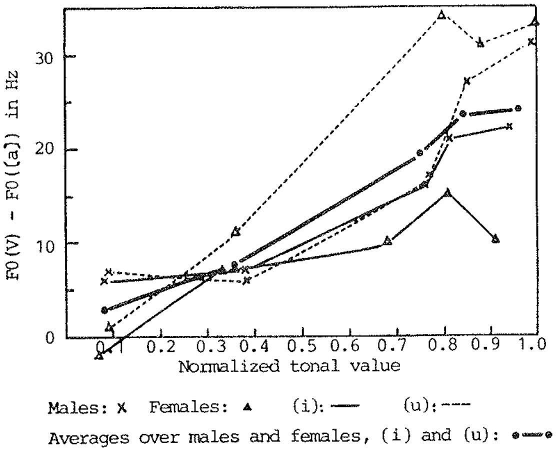

Fig.3 Mean F0 differences between (i,u) and (a) are directly proportional to the normalized tonal value

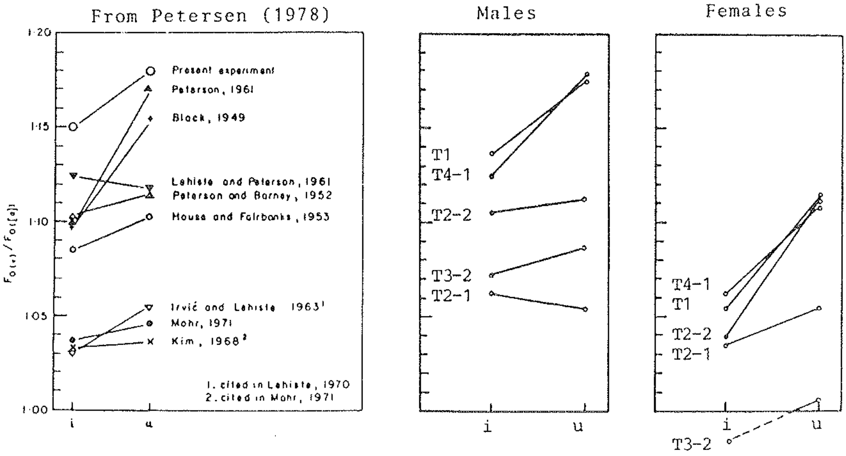

Fig.4 Mean ratios of (i) to (a) and (u) to (a), based on the data from the present and a number of other investigations

questionable.

There is a source/tract coupling hypothesis (See i.e. Flanagan and Landgraf, 1968; Atkinson, 1973) based on the assumption that the IP could be caused by acoustic interaction between the first vowel formant and the vibration of the vocal folds. Since higher vowels have lower first formants than low vowels, the acoustic interaction should be greatest for high vowels whose first formant frequencies are closer in frequency to F0. However, the results of the experiments made by Beil (1962) and Ewan (1979) contradict this hypothesis. And the relationship between [i] and [u] does not tally with what the hypothesis predicts either. We usually find that the first formant F1 of [i] is lower than that of [u]. According to the hypothesis [i] should be expected to have a higher F0 than [u]. Nevertheless, the present and a number of other investigations show the opposite relation, i.e. a higher F0 in [u] than [i] (See Fig.4).

The [i]-[u] relation we just discussed also contradicts another hypothesis -- Mohr's "pressure hypothesis" (1971). Mohr assumes a pressure of air to be built up behind the supraglottal constriction, which reduces the airflow through the glottis and, consequently, the rate of vocal fold vibration. The greater the distance between the constriction and the glottis, the longer it takes for the air pressure to be established behind the constriction and hence for the F0 to drop. If this explanation were true, a higher F0 should be expected in [i] than in [u]. Thus, the results of the present study support neither source/tract coupling nor pressure hypotheses.

Of the various hypotheses, it seems that the tongue pull theory has received the greatest attention. The early tongue pull hypothesis (Ladedoged, 1964) supposed that the tongue, when raised to produce high vowels, pulls the hyoid bone and the larynx upwards, thus resulting in an increased vocal-fold tension which in turn leads to a higher F0. But this explanation is contradicted by the fact that the hyoid/larynx position always seems to be lower in [u] than in [a]. Ohala (1973) modified the tongue pull hypothesis. He thought that the increased tongue pull in high vowels gives rise to increased vertical tension in the vocal folds through the mucous membrane and other soft tissues without involving the hyoid bone and the hard tissues of the larynx. In support of this explanation, it appears that there is a positive correlation between ventricle size, which is assumed to reflect vertical tension in the vocal folds and tongue height and intrinsic F0 of vowels. The tongue pull hypothesis has been expanded further by Ewan (1975). In addition to tongue pull--or rather, lack of tongue pull-- Ewan proposes a "tongue retraction/pharyngeal constriction component" to account for the low intrinsic F0 of low vowels. Ewan suggests that the low F0 of low vowels, which are also assumed to involve a tongue retraction or pharyngeal constriction component, is caused by the soft tissues being pressed downwards in the direction of the larynx and thus increasing the vibrating mass of the vocal folds, which results in a decrease in F0.

But few of these hypotheses attempt an explanation of the "nonlinearity" in IP. In Chinese the higher the tonal value, the larger the IP difference; in Italian, the accented syllables display greater IP than unaccented ones (Ferrero et al., 1975); deaf speakers often exhibit a larger than normal IP which may be related to a higher than normal average F0 (Bush, 1981). IP is reduced in final sentence position with a lowered F0 (Shadle, 1985). The common point is that a larger IP difference seems always correlated to a higher F0. Moreover, the variation of tonal characteristics due to syntactic and semantic factors is much larger at the tonal roof than at the tonal floor (Bannert, 1984). So a larger variation of F0 always corresponds to a higher F0. And this sort of nonlinearity is relative to a <u>within</u>-subject variation (i.e. it does not mean the female should be expected to have a larger IP difference than the male because of a higher voice). There was a simpler explanation that general relaxation (as in an unaccented phrase-final position) may reduce intrinsic F0. But it is contradicted by the evidence against vowel neutralization in that "relaxed" sentence position (Shadle, 1985).

Here, we try to give a probable interpretation from the point of inherent nonlinearity of the vocalis muscle itself. According to Ohala's theory (1973) the tongue pull gives rise to increased vertical tension in the vocal folds through the mucous membrane and other soft tissues. We could assume that there must be a series of deformations in the mucous membrane and the soft tissues, and finally in the vocalis muscle itself thus causing increased tension. The relationship between the tension T and the elongation x of the vocalis muscle can be approximately expressed as (Fujisaki et al., 1981):

$$T=a\ \exp(bx)$$

The incremental tension per unit elongation, as given by $\gamma T/\gamma x$, is obviously greater at larger values of x which generally correspond to higher F0-values. In other words, the same incremental elongation due to the tongue pull could cause a larger increase in tension T, thus leading to a larger F0 variance at high F0 than at low F0. However, it must be emphasized that this is only a probable conjecture. The reliable evidence for the interpretation should be based on physiological data. Last, we think that if this kind of nonlinearity in the production of speech could be confirmed, it would be helpful for a better understanding of the similar nonlinearity found in the perception of speech.

ACKNOWLEDGMENTS

We would like to express our thanks to Prof. Eva Gårding and Prof. Maa Dah-You for their valuable suggestions and critical reading of the early versions. We are also greatly indebted to many colleagues for assistance, for encouragement, and for guiding us to relevant literature: David House, Li Meng, Li

Zi-yin, Lü Shi-nan, Qi Shi-qian, Jan-Olof Svantesson, and Paul Touati.

REFERENCES

Appelman, D. R. 1967. The Science of vocal pedagogy (Indiana U. P., Bloomington).

Atkinson, J. E. 1973. Aspects of intonation in speech: Implications from an experimental study of fundamental frequency. PhD dissertation, Univ. of Connecticut, Storrs, CT (unpublished).

Bannert, R. 1984. Towards a model for German prosody. Working Papers 27: 1-36. Dept. of Linguistics and Phonetics, Lund Univ..

Beil, R.G. 1962. Frequency analysis of vowels produced in a helium-rich atmosphere. J. Acoust. Soc. Am. 34: 347-349.

Bush, M. 1981. Vowel articulation and laryngeal control in the speech of the deaf. PhD thesis, MIT (unpublished).

Connel, B. A. et al. 1983. Experimental evidence of interaction between tone and intonation in Mandarin Chinese. J. Phonetics 11: 337-351.

Crandall, J. B. 1925. The sounds of speech, Bell Syst. Tech. J. 4: 586-626.

Di Cristo, A. 1985. De la microprosodie a l'intonosyntaxe. Publications Université de Provence.

Ewan, W.G. 1975. Explaining the intrinsic pitch of vowels. Paper presented at the Fifth California Linguistics Association Conference, San Jose, May 4, 1975, 1-9.

Ewan, W. G. 1979. Can intrinsic vowel F0 be explained by source/tract coupling? J. Acoust. Soc. Am. 66: 358-326,

Ferrero, F. E. et al. 1975. Some acoustic and perceptual characteristics of the Italian vowels. Paper presented at the VIIIth Int. Congr. Phon. Sci. (Leeds), mimeographed.

Flanagan, J. and Landgraf, L. 1968. Self-oscillating source for vocal-tract synthesizers. IEEE Transactions on Audio and Electroacoustics AU-16, 57-64.

Fujisaki, H et al. 1981. Analysis of pitch control in singing. In K.N. Stevens and M.Hirano (eds.), Vocal fold physiology. Univ. of Tokyo press, 347-363.

Hombert. J-M. 1977. Consonant types, vowel height and tone in Yoruba. Stud. Afri. Linguistics 8: 173-190.

House, A. and Fairbanks, G. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. J. Acoust. Soc. Am. 35: 84-92.

Ivic, P. and Lehiste, I. 1963. Prilozi ispitivanju fonetske i fonološke prirode akcenata u savremenom srpskohrvatskom književnom jeziku. Zbornik za filologijy i lingvistiku 6: 33-73.

Ladd, D. R. and Silverman, K.E.A. 1984. Vowel intrinsic pitch in connected speech. Phonetica 41: 31-40.

Ladefoged, P. 1964. A phonetic study of West African languages. In West African Language Monograph, Series I (Cambridge U. P., Cambridge).

Lehiste, I. and Peterson, G. E. 1961. Some basic considerations in the analysis of intonation. J. Acoust. Soc. Am. 33: 419-425.

Lubker, J. et al. 1977. Vowel fundamental frequency and tongue height. J. Acoust. soc. Am. 62: S16-17(A).

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

Neweklowsky, G. 1975. Spezifische dauer und spezifische tonhöhe der vokale. Phonetica, 32, 1: 38-60.

Nishinuma, Y. 1977. Contribution à l'étude prosodique du Japonais. Thèse de 3e Cycle. Univ. de Provence.

Ohala, J. 1973. Explanations for the intrinsic pitch of vowels. Monthly Internal Memorandum, Phonology Laboratory, Univ. of California, Berkeley (January), 9-26.

Petersen, N. R. 1976. Intrinsic fundamental frequency of Danish vowels. J. of Phonetics (1978) 6: 177-189.

Peterson, G. E., and Barney, H. L. 1952. Control methods used in a study of the vowels. J. Acoust. Soc. Am. 24: 118-127.

Samaras, M. 1972. Influence de l'entourage consonantique sur les variations de la fréquence laryngienne des voyelles du grec moderne, Bull. de l'Inst. de Phon. de Grenoble, I: 57-66.

Shadle, C. H. 1985. Intrinsic fundamental frequency of vowels in sentence context. J. Acoust. Soc. Am. 78: 1562-1567.

Taylor, H. C. 1933. The fundamental pitch of English vowels. J. Exp. Psychol., 16: 565-582.

Umeda, N. 1981. Influence of segmental factors on fundamental frequency in fluent speech. J. Acoust. Soc. Am. 70: 350-355.

Zee, E. 1978. The interaction of tone and vowel quality. UCLA work. Pap. Phonet. 41: 53-67.