VOWEL INTRINSIC PITCH IN STANDARD CHINESE

SHI BO and ZHANG JIALU

Institute of Acoustics
Academia Sinica
Beijing, China

ABSTRACT

We investigated whether an intrinsic pitch (IP) effect occurs in Standard Chinese 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 cases: (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 four different tones of Standard Chinese in spite of the fact that those tones have different F0-patterns. 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 (or 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, Italian, Danish, Japanese, French, German, Greece, Taiwanese Chinese, Yoruba, Serbo-Croatian, Itsekiri, and Chinese. IP has also been observed when vowels were sung at the same pitch. The reference list can be found in [1].

Various experimental conditions were applied in these studies. In the early experiments, isolated 'real' words as well as 'nonsense' words were used. The segmental environments (i.e. consonantal context) were carefully controlled. Later, the test words were embedded in a frame sentence. The effects of prosodic environment on IP had been took into account. Petersen [2] reported that the magnitude of IP in stressed syllables is larger than the one in unstressed syllables. Similar results were obtained for Italian accent/nonaccent words [3]. All of these studies generally showed similar results except Umeda's [4] 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 speech, Ladd and Silverman [5] 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 [6] 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 .

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. All possible combinations of consonants and simple vowels in Standard Chinese were included in the monosyllable part, and each combination occurs four times with four different tone patterns. Among them there are 279 'real' monosyllabic words and 121 'nonsense' words. In the disyllabic word part, every word consist of one test syllable (a simple 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 fǎhuà/fǔhuà; wěibā/wěibō/wěibī; tǔjìng/tǐxìng, (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 be in the same phonetic environment and to 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 gè cí./ (I utter the word __ __.) respectively.

Ten speakers (5 males and 5 females) of Standard Chinese were recorded. They had been trained for a short period before the recordings. A natural speech style was aimed at. The test materials were read once by each speaker in an acoustically treated room.

The recordings were fed into a Visi-Pitch (model 6087) for the extraction of F0. The counter on the Visi-Pitch 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.

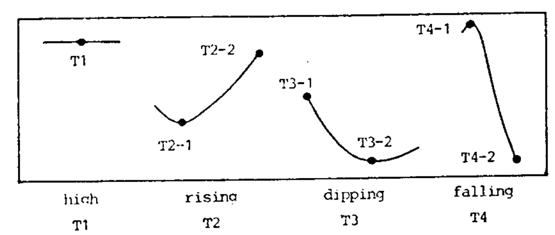


Fig.1 Measuring points of fundamental frequency

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

RESULTS

1. Vowels Intrinsic Pitch In Four Tones

The data which will be analysed in this section were derived from 400 monosyllables. The intrinsic F0-values for each of 9 vowels and relative F0 differences (ΔFO) between the vowel [a] and the remaining 8 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. This is also shown graphically in Fig.2 (see •—•).

The data mentioned above permit us to make the following observations: 1) at points T1, T2-2, and T4-1, the F0-values of the vowels go from high to low as the tongue height of the associated vowel drops, and the F0 differences between high and low vowels are significant; 2) at points T2-1 and T3-2 a high vowel also has a higher F0 except that the F0-value of [o] of the males is a bit higher than that of [ʅ] and [i] and the FO-value of [a] of the females is higher than what is expected. The data at these five points show that Chinese, as a tone language, also exhibits the influence of intrinsic pitch.

The situation is more complex at points T3-1 and T4-2. We found considerable inter- and intra-speaker variability for F0-values at point T3-1. The main problem at point T4-2 is that the energy at the end of T4 is very low and the periodicity is not good enough to permit precision in measurements. As a result there is no consistent influence of IP at these two points.

Table 1. Mean intrinsic F0-value for each of the 9 Chinese vowels and relative FO differences (ΔFO) between the vowel [a] and the remaining 8 vowels at different tonal points, derived from 400 monosyllables, averaged across consonantal contexts, and for 5 males and 5 females respectively.

FO and ΔFO (Hz)

	T1		T2-1		T2-2		T3-1		T3-2		T4-1		T4-2	
	FO	ΔFO	FO	ΔFO	FO	ΔFO	FO	ΔFO	FO	ΔFO	FO	ΔFO	FO	ΔFO
(5 males)														
i	175	21	118	7	167	16	113	5	89	6	197	22	97	0
ɿ	181	27	122	11	171	20	116	8	90	7	208	33	99	2
ʅ	179	25	116	5	169	18	115	7	90	7	195	20	101	4
y	180	26	119	8	175	24	115	7	90	7	197	22	101	4
u	181	27	117	6	168	17	112	4	90	7	206	31	105	8
e	164	10	114	3	156	5	114	6	88	5	187	12	101	4
o	168	14	117	6	160	9	116	8	90	7	184	9	100	3
ɚ	170	16	116	5	170	19	122	14	88	5	178	3	100	3
a	154	0	111	0	151	0	108	0	83	0	175	0	97	0
(5 females)														
i	291	15	205	7	265	10	219	-8	169	-2	312	10	180	-7
ɿ	302	26	206	8	271	16	214	-13	172	1	326	24	182	-5
ʅ	295	19	200	2	264	9	216	-11	168	-3	319	17	192	5
y	300	24	209	11	278	23	219	-8	171	0	318	16	176	-11
u	307	31	209	11	289	34	218	-9	172	1	335	33	184	-3
e	289	13	202	4	270	15	215	-12	170	-1	315	13	183	-4
o	278	2	200	2	270	15	213	-14	170	-1	310	8	183	-4
ɚ	302	26	200	2	274	19	209	-18	161	-10	314	12	182	-5
a	276	0	198	0	255	0	227	0	171	0	302	0	187	0

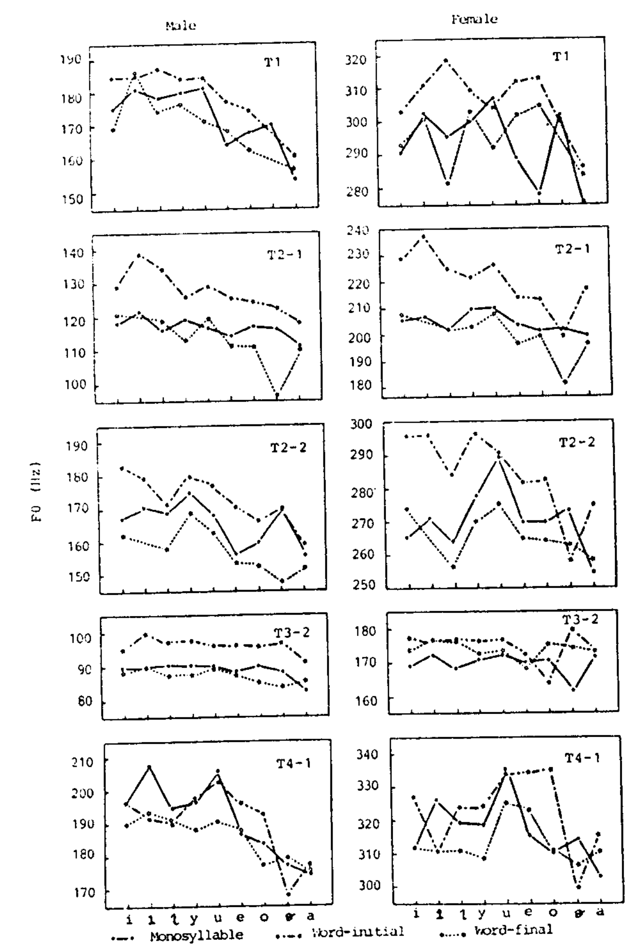


Fig.2 Mean F0 for the vowels, plotted as a function of tongue height, averaged across consonants and for 5 males and 5 females respectively

2. Effect of Word-position on IP

There are additional factors influencing F0-value of the vowels when the test syllables were in the disyllabic words. For instance, the F0-pattern of the test syllable could be modified by the adjacent tones as well as might vary with different stress pattern caused by different semantic meaning.

The data in Fig.2 (⁎and ☆) show that, though semantic 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. So the variation of tonal characteristics due to intrinsic pitch is larger than the one due to semantic and tonal environment factors. 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

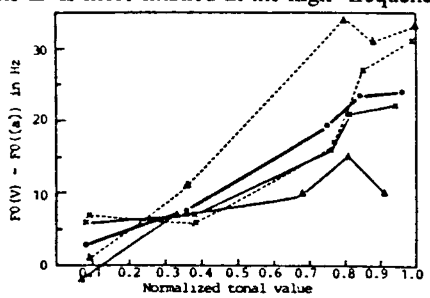
Fig.2 shows F0-values of 9 simple vowels as a function of the tongue height associated with them. Generally speaking, in each part of Fig.2, from left to right, the tongue height of the vowel goes from high to low and it is accompanied by a drop in F0, which reflects the effect of IP. But the curves in Fig.2 at different tonal points have different slopes, i.e. the differences of intrinsic F0 (ΔF0) across the vowels vary from point to point (also see Table 1.). The ΔF0 at points T1 and T4-1 (high F0) are obviously much larger than those at 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 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 describe 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 normalised 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.



Males: ⁎ Females: ▲ (i): — (u): ---

Averages over males and females, (i) and (u): •—•

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

But the slopes of the two curves of the females turn negative when the normalised 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 is might be worth while to study intrinsic pitch in a larger F0 dynamic range such as in singing.

DISCUSSION

There have been various hypotheses for the cause of IP: dynamogenetic irradiation hypothesis[7], source/tract coupling hypothesis [8], pressure hypothesis [9], and tongue pull hypothesis.

Of the various hypotheses, it seems that the tongue pull theory has received the greatest attention. The early tongue pull hypothesis [10] 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 [11] modified the tongue pull hypothesis. He thought that the 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 [12]. 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 [3]; deaf speakers often exhibit a larger than normal IP which may be related to a higher than normal average F0 [13]. IP is reduced in final sentence position with a lowered F0 [6]. The common point of these results 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 [14]. So a larger variation of F0 always corresponds to a higher F0. And this sort of nonlinearity is relative to a within- 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 phrased-final position) may reduce intrinsic F0. But it is contradicted by the evidence against vowel neutralization in that 'relaxed' sentence position [6].

Here, we try to give a probable interpretation from the point of inherent nonlinearity of the vocalis muscle itself. According to Ohala's theory 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:

$$T=ae^{bx}$$

The incremental tension per unit elongation, as given by $\partial T/\partial x(=abe^{bx})$ 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.

ACKNOWLEDGEMENTS

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.

REFERENCES

[1] Shi, Bo and Zhang Jialu. 1986. Vowel intrinsic pitch in Standard Chinese. Working Papers 29: 169-190. Dept. of Linguistics and Phonetics. lund Univ..

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

[3] 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.

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

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

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

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

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

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

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

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

[12] 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.

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

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