THE INTRINSIC FUNDAMENTAL FREQUENCY OF VOWELS AND
THE EFFECT OF SPEECH MODES ON FORMANTS

JIALU ZHANG

Institute of Acoustics, Academia Sinica
Beijing, CHINA

ABSTRACT

Two experiments have been carried out to explore the interaction phenomena. It is shown that 1) The intrinsic fundamental frequency of vowels is also found in Chinese, 2) The difference of intrinsic fundamental frequency between high and low vowels are related to the pitch level and a linear relationship was found in a certain dynamic range of register beyond that nonlinear relation will appear, 3) speech efforts influence not only the F0 but also the F1-F2 pattern of vowels.

INTRODUCTION

In recent years a more profound view of the speech production process is emerging and interactive models taking into account the interaction between source and vocal tract have been proposed [1,2,3,4]. In these models more attention has been paid to the acoustic interaction. However some investigations [5] shown that the mechanical interaction is prominent in some instances.

We still need, however, to incorporate more knowledge of source-filter interaction and speaker/speaking particulars to improve the model of speech production. In order to get an insight into the model, we have to get much more experimental data in dynamic process of connected speech.

The present paper deals with two interaction phenomena. The first concerns the relation between intrinsic fundamental frequency (henceforth IF0) for vowels in different tonal environments and different syllable structures, with groups of both adult male and female speakers. The second source of interaction studied comes from the dependent of vowel formant frequencies on speech modes.

EXPERIMENT I

A great deal of research has been devoted to the analysis and quantification of IF0 in several languages. However, none of these studies were concerned with the roles of pitch level and the syllable structures and word position as determinants of IF0. The speech material used in this study consists of two parts, 400 monosyllables and 509 disyllabic words.

In order to make all test items occur in the same phonetic environment and approch the situation of connected speech, all the monosyllables and disyllabic words were embedded in a frame sentence "Wǒ dú ___ zì." (I utter the character ___.) and "Wǒ dú ______ zhège cí." (I utter the word ______.) respectively. Ten speakers (5 male and 5 female) who speak "Putonghua" (standard Chinese) were recorded.

The measuring points of F0 are at the middle point for level tone T1; at lowest point T2-1 and highest point T2-2 for rising tone T2; at starting point T3-1 and lowest point T3-2 for dipping tone T3 (because in connected speech the tone contour of T3 will change from falling and rising into falling and low level except T3 is followed by another T3); at highest point T4-1 and lowest point T4-2 for falling tone T4.

The mean IF0 for each of nine Chinese vowels at different tonal points, and the IF0 difference between other vowels and /a/, ΔF0, derived from 400 monosyllables, averaged across consonantal contexts, and for 5 male and 5 female speakers respectively are listed in Table 1.

Table 1. Mean IF0 and IF0 difference between other vowels and /a/, ΔF0, at different tonal points for 5 males and 5 females respectively.

F0 and F0, (Hz), 5 males

	T1		T2-1		T2-2		T3-1		T3-2		T4-1		T4-2	
	F0	ΔF0	F0	ΔF0	F0	ΔF0	F0	ΔF0	F0	ΔF0	F0	ΔF0	F0	ΔF0
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

	T1		T2-1		T2-2		T3-1		T3-2		T4-1		T4-2	
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

From Table 1. it can be seen that Chinese, which is a language with multitone system, also exhibits vowel IF0. Some negative values of F0 appeared at tonal points T3-1, T3-2, and T4-1 for females, this is due to the problems of F0 extraction. And it is worth to note that: 1) The IF0 difference between high vowels and low vowel /a/, F0, are related to the vowel pitch level; 2) There are no significant differences in the values of F0 between males and females. The IF0 difference, F0, is mainly dependent on the tonal value. The relation between F0 and pitch level is shown in Fig. 1. Two different kinds of tonal scale were used as abscissa, one is relative or normalized tonal value, which is defined as the average tonal value to tonal register ratio; the other one the average value of F0 corresponding to these tonal values over all speakers, both male and female.

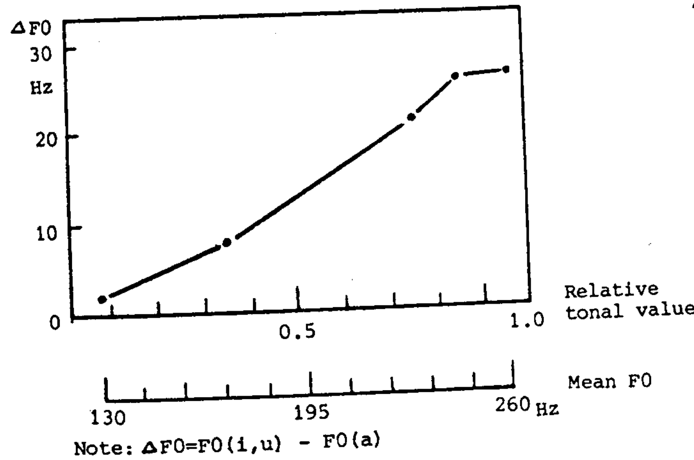


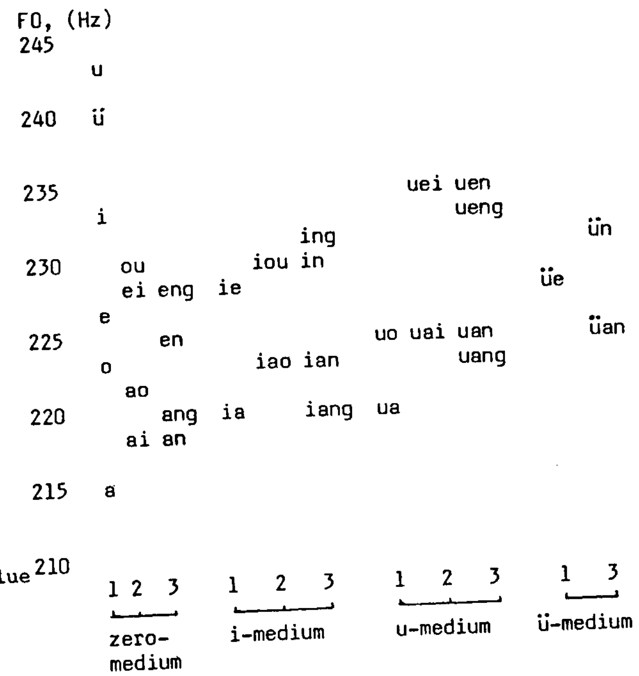
Fig. 1. ΔF0 versus mean F0 for male and female speakers at different tonal points.

From Fig. 1. it can be seen that F0 increases linearly with F0 for about an octave and after the linear section some nonlinearity or saturation of ΔF0 appears. Some similar phonomena have been observed in Italian where the accented syllables displaying greater IF0 than unaccented ones [6]. IF0 is redused in the final sentence positions with a lowered F0 [7]. In summary, a larger IF0 difference is generally related to a higher F0 (tonal value).

In order to show the influence of vowel combinations and of vowel nasalizations on the IF0, the F0 of "Yùnmǔ" (finals), with tone T1 averaged across consonantal contexts for 5 male and 5 female speakers, are drawn in Fig. 2. according to their structures.

Fig.2. shows that: 1) The nasalization of vowels reduce the IF0 difference between high vowels and low vowels, some similar results were found in French [5]; 2) "Yùnwěi" /i,u,ü/ (vowel final endings) tend to increase the IF0 of main vowels; 3) "Jiemǔ" (medium vowels) /i,u,ü/ increase the IF0 of main vowels and main vowels with final endings (both vowel endings and nasal endings). In other words, diphthongs and triphthongs tend to reduce the IF0 difference. This is perhaps due to the fact that the supraglottal configuration is changed continuously and quickly during the pronunciation of compound vowels and the mechanical interaction has a certain time constant (due to the inertia of the muscles) so that F0, which is a laryngeal parameter, can not be changed as fast as the supraglottal tongue movement.

There have been various hypotheses concerned with the cause of vowel IF0. But none of them take notice of the "linearity" and "nonlinearity" of



Note: (1) /ao,iao/ are actually /au,iau/ in phonetic value.
(2) The number 1, 2, and 3 stand for the "yùnmǔ" without endings, with vowel endings, and with nasal endings in the corresponding column respectively.

Fig. 2. The influence of vowel combinations and vowel nasalizations on vowel IF0.

vowel IF0 in relation to larynx frequency. It seems that the tongue pull [8] or mechanical interaction theory has greater significance than other propositions. Here I try to give a probable interpretation from the point of mechanism 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 corresponding structural change in the mucous membrane and the soft tissues, and finally in the vocalis muscle itself thus causing a increased tension. The relationship between the tension T and the elongation X of the vocalis muscle can be approximated as [9]:

$$T=aEXP(bx) \quad (1)$$

and to a first-order approximation the fundamental frequency F0 of the vibration of vocal folds as

$$F0=c0\sqrt{T} \quad (2)$$

Then the incremental tension per unit elongation can be expressed as

$\partial T/\partial X=bT$ (3)

If we neglect that c0 varies slightly with X, then we obtain

$\partial F0/\partial X=(1/2)bF0$ (4)

Formula (4) shows a linear relationship between $\Delta F0$ and F0, in other words, the same incremental elongation ΔX due to tongue pull could cause a larger increment in tension ΔT, thus leading to a larger increment of fundamental frequency $\Delta F0$, at high F0 than low F0. However, it seems that there is a dynamic range (about one octave in our data) for speakers control of their vocal folds. Beyond the dynamic range "nonlinearity" appears, perhaps, parameters a,b, and c0 in formulas (1) and (2) are not constant and the vibration of the vocal folds is associated with "overloading" near the upper end of the register.

EXPERIMENT II

Generally, it is assumed that formant frequencies are entirely determined by the vocal tract and independent from the voice source. And a target specified by the formant frequency of a vowel was considered as an invariant attribute. Recently some experimental results on dynamic spectra of speech sounds show that speech efforts and speech speed give strong influence not only on F0 but also on the F-pattern [10]. The testing material is a well designed sentence "Tā qù Wúxīshì, wǒ dào Hēilóngjiāng."(He comes to Wuxishi, I go to Heilongjiang.), in which three primary vowels /i,u,a/ are included. 12 speakers (6 male and 6 female) who are natives of Beijing uttered the testing sentence repeatedly in different speech modes in an anechoic chamber. The speakers were asked to change their speech efforts in five levels: 55, 60, 65, 70, and 75 dB SPL measured at 1 meter in front of the speaker's lips and to speed up the speaking rate by a factor of 2 at 65 dB SPL.

The F1-F2 plane for speech levels 55, 65, 75 dB and the high speed version are drawn in Fig.3. The variation of F0 with the SPL of speech for three syllables with different pitch levels are shown in Fig. 4.

Table 1. The average increments of F1 and F2, $\Delta F1$ and $\Delta F2$, for different vowels from 55 to 75 dB SPL of speech over 6 male and 6 female speakers.

	i	ü	u	-i	a
$\Delta F1$, Hz	104	90	82	70	202
$\Delta F2$, Hz	279	158	128	137	230

Note: -i stands for either [ɿ] and [ʅ].

From Fig. 3.a. it can be clearly seen that the vowel triangle is shifted and enlarged regularly as the SPL of speech increases. This suggests that both jaw and tongue movements are enlarged when a talker speaks loudly. On the other hand when the tempo is speeded up all vowels tend to be centralized and the vowel triangle is reduced (see Fig. 3.b.). The increments of F1 and F2 from 55 dB to 75 dB SPL of speech for different vowels are given in Table 1.

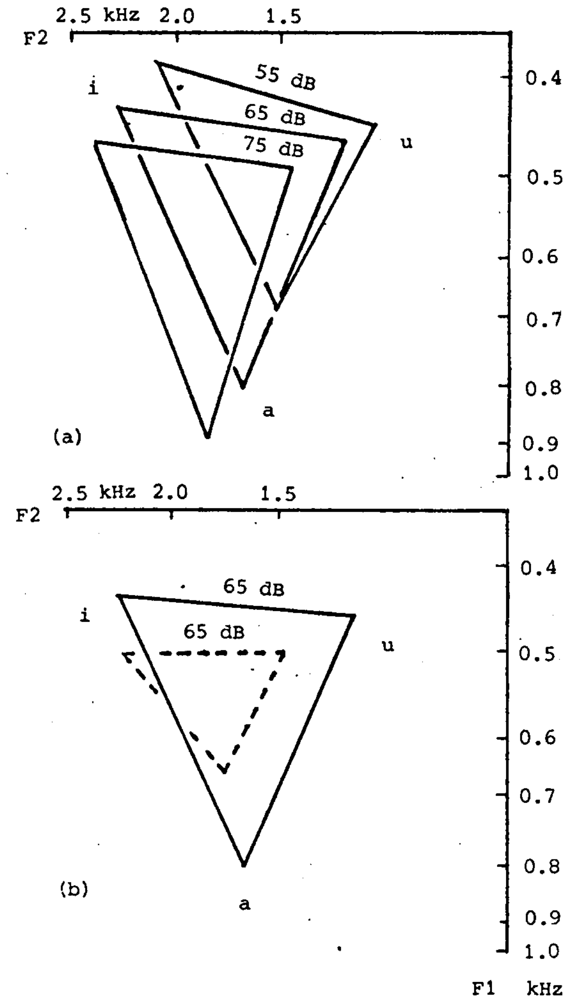


Fig. 3. F1-F2 plane for different speech efforts (a) and different utterance speeds (b).

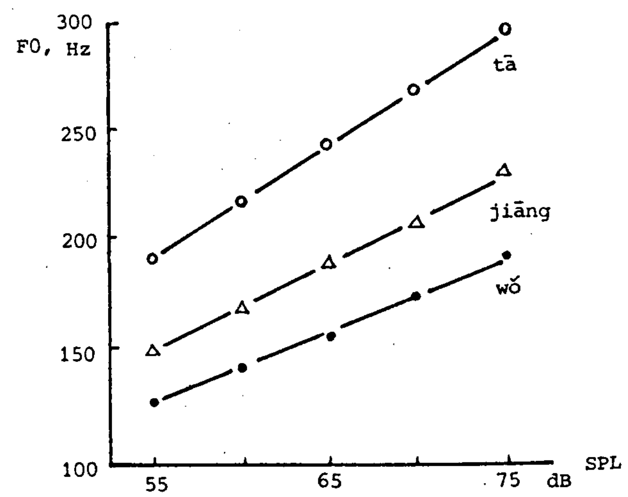


Fig. 4. F0 of main vowels with different pitch levels versus SPL of speech.

As the speakers increase their effort F1 shifts up in frequency by about 100 Hz for all vowels except /a/ which is raised by about 200 Hz. F2 shifts up in frequency somewhat more than F1. As for F3 the variations of all vowels are irregular.

Fig. 4 shows that the F0 increases linearly with the SPL of speech but that the steepness varies with the pitch levels of the vowels. The syllable "Tā" with level tone T1 in initial sentence position has the highest pitch level and the largest slope in F0 with increasing SPL of speech. The syllable "Wǒ" with dipping tone T3 has the lowest pitch level in the sentence and the smllest slope. As for the syllable "jiāng" with T1 in the sentence final position, it gets a mid pitch level and a middle slope in F0 because of declination. The average increment of F0 with increasing SPL of speech from 55 dB to 75 dB over 6 male and 6 female speakers is about 111 Hz for high pitch vowel, about 98 Hz for mid pitch vowel and about 73 Hz for low pitch vowel. However, the relative increment of F0 or the increment in log F0 scale for the three syllables are nearly the same about 60 per cent i.e. F0(75dB)=1.6F0(55dB).

According to Fant [1] there is a net gain of 9.5 dB included all factors accompanying the doubling of lung pressure. Starting out from F0=149 Hz, SPL=55 dB, say lung pressure Pl=5 cm H2O and increasing SPL of speech to 75 dB, an increase of Pl should be to 20 cm H2O. And the increase of Pl by 15 cm H2O cause an increment of F0 by 73 Hz for low pitch vowel and by 111 Hz for high pitch vowel, in other words the incremental rate of F0 covers from 4.8 to 7.4 Hz/cm H2O. That is a reasonable value, but somewhat higher than the predicted 3.5 Hz/cm H2O [11].

Comparing the increments of F0 of vowels of different pitch levels caused by increasing lung pressure with the IF0 difference related to F0 of vowels, some similarity appears which might have a common physiological basis.

So when a speaker increases his/her lung pressure during speaking, both F0---source parameter and F1,F2---vocal tract parameters are simultaneously increased. This is due to the increased air push force which makes the glottis shift upward, then the tension of vocalis muscle is increased and the lenghth of vocal tract is shortened. This effect adds to increasing mouth opening. We searched persistently an invariant attribute of vowel targets but we found some floating islands---moving vowel triangles instead. It tells us that relative position of vowels in perceptual space are very important.

CONCLUSION

Two different experiments have illustrated interactions of phonatory and articulatory mechanisms. It seems from the present results that the mechanical interaction is stronger than acoustic interaction under these conditions. A lot of experimental results show that the intrinsic fundamental frequency of vowels is universal and we find here that it is also exhibited in Chinese---a tone language with multitone system. The intrinsic fundamental frequency difference $\Delta F0$ increases linearly with pitch level over a certain dynamic range and then saturates in a nonlinear region near the upper boundary of the tonal range. The tongue pull theory and the stress-strain relationship of muscle could account almost entirely for vowel IF0.

Speech effort not only influences the F0 but also the F1-F2 plane. It is hard to determine vowel targets in connected speech, and the vowel triangle is floating. The relative positions of vowels in the F1-F2 plane, however, convey the perceptual features of vowel quality and relate to the speech modes. The fact that F0 increments caused by speech efforts are related to pitch level is somewhat similar to the IF0 difference caused by tongue pull. And a subglottal air push hypothesis could explain the fact that both F0 and F1-F2 plane pattern are changed with speech effort.

There remains much work to be done on the theoretical modeling and in regard to the development of experimental techniques to establish an advanced interactive model of speech production.

ACKNOLEDGEMENTS

The author wishes to thank Professor Fant for his very valuable comments and enthusiastic encouragement in this work and to express his sincere appreciation of Professor Fourcin's help.

REFERENCES

[1] Fant, G. (1982A): "Preliminaries to analysis of the human voice source", STL-QPSR 4/1982, 1-28.
[2] Fant, G. (1982B): "The voice source---acoustic modling", STL-QPSR 4/1982, 29-49.
[3] Fant, G (1985): "A four-parameter model of glottal flow", STL-QPSR 4/1985, 1-13.
[4] Rothenberg, M. (1981): "An interactive model for the voice source", STL-QPSR 4/1981, 1-17.
[5] Delos, M. et al. (1976): "Study of intrinsic pitch of vowels", J. Acoust. Soc. Am. 59,S72(A).
[6] Ferrero, F. E. et al., (1975): "Some acoustic and perceptual characteristics of the Italian vowels", VIIIth Int. Cong. Phon. Sc., (Leeds),mimeographed.
[7] Shadle, C. H. (1985): "Intrinsic fundamental frequency of vowels in sentence context", J. Acoust. Soc. Am., 78, 1562-1567.
[8] Ohala, J. (1973): "Explanations for the intrinsic pitch of vowels", Monthly Internal Memorandum, Phonology Lab.,Univ. of California, Berkeley (January), 9-26.
[9] Fujisaki, H. et al. (1981): Analysis of pitch control in singing", in Vocal folds phsyology, K. N. Stevens and M. Hirano (eds), 347-363, Univ. of Tokyo Press.
[10] Zhang, Jialu and Qi, Shiqian (1982): "On the dynamic spectra of speech", FASE/DAGA'82, 997-1000.
[11] Fant, G. and J. Liljencrants (1979): "Perception of vowels with truncated intraperiod decay envelops", STL-QPSR 1/1979, 79-89.