INTRINSIC VOWEL DURATION IN SHANGHAI OPEN SYLLABLES

Xiaonong Sean Zhu

Linguistics Department, Faculties
Australian National University

ABSTRACT - There is a contradiction between the observed phenomenon that low vowels tend to be longer than high vowels (the Lehiste IVD) and the theoretical prediction that high vowels should be longer than low vowels (the Catford IVD). This paper argues that the Lehiste IVD is environmental conditioned by a following consonant. The experiments, using the Shanghai dialect, conducted to test the IVD show that the Catford IVD varies with F0 directions: 1) in a CV syllable with a falling tone, a high vowel is longer than a low one; and 2) in a CV syllable with a rising tone, high and low vowels have comparable duration.

INTRODUCTION

It has long been observed in many languages that duration varies with reference to vowel height; other factors being equal, low vowels such as [a] tend to be longer than high vowels such as [i, u] (cf Lehiste 1970:18 ff). Let us call 'a>i' ([a] being longer than [i]) the Lehiste version of intrinsic vowel duration (IVD).

There have been two types of hypotheses to account for Lehiste IVD. The first implicates mechanical movement (e.g. Lindblom (1967) and Clark and Yallop (1990:54)). Condax and Krones (1976) find that the duration differences between high and low vowels, considered to be attributable to the mechanical constraints, disappear when subjects manually control synthetic vowel duration, whilst that between paired tense and lax vowels, which is thought to be learnt and programmed in the brain as part of grammar, remains.

The other type of hypotheses invokes neural control. Keating (1985) reports an EMG experiment conducted by Westbury and Keating and claims that the longer durations of low vowels were due to longer travel times, not longer steady states. They also found that longer and more extensive movements were accomplished with a force input that was both longer in duration and higher in amplitude. This led them to the conclusion that the vowel duration differences are not caused by sluggishness of jaw, but rather are centrally controlled as such. They thus predict there should be languages with different vowel-duration patterns. For example, a low vowel, if it is phonologically short, will be uttered with high speed and thus is shorter than a phonologically long high vowel.

There are two points needed to be made concerning Westbury and Keating's conclusion. First, observable muscle activities do not necessarily imply voluntary control. Their claim holds only when a contrastive experiment has been done with regard to extrinsic vowel duration. Second, it not only does not explain why some vowels are longer than others, it even tends to predict the opposite would happen: if duration was centrally controlled, we would expect that it would be identical for all vowels, unless there had been other such as phonological contrast and phonetic environment. This then leads us to return to the mechanical hypothesis.

It is intriguing to note the opposite theoretical prediction: It should be high rather than low vowels that last longer. Catford (1977:45) suggests that, just like the exhalation time for [s] is longer than that for [θ] due to the fact that the articulatory channel area of [s] is smaller than that of [θ], the low vowels with narrower articulatory channels 'will readily be seen' to 'have longer exhalation times'. The assumption underlying his proposal is that the initial pressure, the additional efforts for pressure, if any, during the course of speech, and the total volume velocity are identical for the productions of [i] and [a]. From this viewpoint follow these inferences. Since the constriction for [i] is smaller than that for [a], hence larger impedance for [i]. A larger impedance will induce smaller flow if pressure keeps constant, as volume velocity is proportional to pressure and inversely proportional to impedance. And the smaller flow will in turn cause less loss of the initial pressure and sustain the phonation longer. This results in a prediction that the duration of high vowels should be longer than that of low ones, allowing the intra-oral pressure, if speakers do not deliberately adjust their musculature involved (no central control). This can be called the Catford version of IVD: 'i>a'.

The contradiction between the theoretical prediction (the Catford IVD) and the experimental observations (the Lehiste IVD) may arise from production contexts. Recall that the test words in English are the type of CVC syllables (Peterson and Lehiste 1960, House 1961), of which the last C makes the courses of jaw movement different with respect to the height of previous vowels, and that it is the travel time rather that the steady state that makes vowel-duration differences. Catford's position, on the other hand, states the duration of the vowel proper, i.e. the steady state of the vowel *per se* rather than the transition phase. If the syllable type used for experiments is like that in this study, viz CV, thus there are only steady states, then a longer [i] and shorter [a] should be expected. It appears that the preceding consonants have no, or negligible, influence on the duration of vowels following them (Lehiste 1976:227), so they are orthogonal to this issue; the results in this study have borne out this suggestion.

Let us now turn to the (C)V case in which the high vowels are supposed to be longer than the low ones, according to the Catford IVD. A potential counter-example against this suggestion might come from Thai data consisting of CV syllables. Thai is also cited in Lehiste (1970:18) to support the 'a>i' IVD claim. A re-examination of the original Thai data (Abramson 1960:118-120) shows that the low vowels, geminate or single, are not consistently longer than the high geminate or single vowels, respectively. For example, for informant WN, the mean duration of [á] with a high level tone, 80 ms, shorter than that of both [í], 100 ms, and [ɔ́], 100 ms; the mean duration of the geminate [ɔɔ̌] with a rising tone, 450 ms, shorter than that of both [uu], 520 ms, and [oo], 500 ms.

Voicing depends on the tension of vocal cords as well as the air pressure difference across glottis. The estimated minimal transglottal pressure drop is 4 cm H_2O to initiate voicing (Flanagan 1972:11) and 2-3 cm H_2O to sustain it (Lieberman *et al* 1969). This minimum 2-4 cm H_2O driving pressure for voicing, however, can only vibrate very loose vocal cords at one's low frequency range, i.e. 'at the softest possible level' (Flanagan 1972:11). Along with the increase of F0, the required minimum driving pressure goes up accordingly. Now consider two vowel cases, one with a falling tone towards the low part of one's F0 range, the other with a rising tone towards the high part of one's F0 range.

In the falling tone case, intensity, which is directly related to Ps, decays side by side with F0 (Zhu 1995). Since the volume-velocity for a high vowel, say, [i], is smaller than that for a low vowel, say, [a], the remaining Ps for [a] at the last part of the duration should be lower than that for [i]. Consequently, the Ps for [a] will reach the voicing pressure threshold earlier, that is, the phonation for [a] will cease earlier, than that for [i].

In the rising tone case, intensity drops drastically from the 80% of the duration while F0 is still climbing steadily to reach its peak at the 100% of the duration. One may assume, again, [a] would stop vibration earlier than [i] owing to the same physiological reason put forward for the falling tone. However consider the IF0 effect which occurs especially at the higher part of one's F0 range (Zhu 1992). With a rising tone [i] is significantly higher in Hz than [a] with the same tone at the later portion of the duration. To maintain voicing for more intensive vocal cords, i.e. with greater impedance, requires greater driving pressure. The inference is that the phonation for [i] would cease earlier than that for [a]. Since we still lack direct measurements of the degree of vocal cord tension, we have to be contented with the speculation that [a] may be longer or shorter than, or equal to, [i] in case of (high) rising tone, as the two contradictory factors regarding the lengthening of high and low vowels may offset each other.

The present study is intended to test the following hypotheses: 1) in a CV syllable with a falling tone, a high vowel is expected to be longer than a low one; and 2) in a CV syllable with a rising tone, high and low vowels will possibly have the similar duration.

METHOD

Corpus and Informants. CV syllables with one of the three long tones in Shanghai Chinese were used in this study. Using the five-level scale (with [5] being the highest and [1] the lowest), the three tones can be represented as: falling [51] for T1, high rising [34] for T2, and low rising [14] for T3. All the syllables used for experiment consist of [i] or [a] preceded by a voiceless unaspirated [p]. There is audible breathy phonation accompanying T3 syllables. The corpus comprises three minimal pairs differing only in vowel height: [pì] 'edge' vs [pà] 'dad' (with T1), [pi] 'flat' vs [pa] 'to put' (T2), and [p̈i̤] 'convenient' vs [p̈a̤] 'to fail' (T3). At the end of T2 and T3 syllables there is a non-phonemic glottal stop, which is reflexed on narrow-band spectrograms as a sharp falling tracing after the F0 peak. Eleven Shanghai native speakers were recruited, aged from 27 to 43. Six of them were men and five women, respectively coded as M1, M2, ..., and W1, W2, ..., and so on.

Recording. The recording was made in a sound treated booth at a single session for each speaker, using a Nagra 4.2 reel-to-reel tape recorder and Nakamichi CM300 cardioid microphone. Sound pressure level was manually adjusted at the beginning of each session. The test words were written in Chinese characters on two pieces of A4 size paper with distracting words before, between, and after them. Informants were instructed to read the word list in normal manner. The word list was read six times with brief pause between each reading. There were thus 66 tokens (11 speakers X 6 repetitions) for each test word.

Measurement. Durational data were obtained from spectrograms made on a Sonograph. Each spectrogram contains two synchronised pictures: a wide-band and a narrow-band. The vowel onset (the 0% timepoint) was decided at the first striation of F1 on the wide-band spectrograms. The offset of vowels with T2 or T3 was decided on narrow-band spectrograms at the peak of the fundamental, ignoring the sudden drop after the peak which is caused by the syllable final glottal stop. Two sets of offset were decided for vowels with T1, one from the consideration of linguistic point of view (marked as 100% below) and the other from the physiological and acoustic grounds (100+%). The 100% point was decided on wide-band spectrograms where the irregular intervals between striations appeared, and the 100+% point was traced to the very end of any discernible tracings of the fundamentals on narrow-band spectrograms.

RESULTS AND DISCUSSIONS

Testing the first hypothesis. The durational data in ms for the falling tone (T1), from the 0% to 100% timepoints, are given in the top half of Table 1. The durational differences between high and low vowels can be read off in the '[i]-[a]' row. The probabilities from a series of 2-tail *t*-tests are given in the bottom half of the table (M stands for the men group, W for women).

TABLE 1. Mean durations and SDs in ms for [i] and [a] with the falling tone (0%-100%)

T1		W2	W5	M6	M2	W3	W1	M5	M1	W4	M3	M4
	n											
[i]	*m*	209	208	247	250	187	310	250	198	255	214	278
	6											
	s	14	25	40	15	16	24	31	26	50	22	27
[a]	*m*	245	234	247	247	184	295	217	154	202	147	194
	6											
	s	27	15	25	24	18	54	30	8	44	28	24
[i]-[a]		-36	-26	0	3	3	15	33	44	53	67	84

2-tail		[i]	[a]	*n*	[i]-[a]	*p*	DF
T1	M	239 (37)	201 (46)	36	39	***	70
	W	234 (52)	232 (51)	30	2	~	58
	All	237 (44)	215 (50)	66	22	**	130

p (***) < 0.001 ² *p* (**) < 0.01 ² *p* (*) < 0.05 ² *p* (~)

The duration of [i] is longer than that of [a] for the male group by 39 ms which is highly significant ($p<0.001$). It is only 2 ms longer for W however. Overall, [i] is longer than [a] by 22 ms and this is significant at 0.01 level. This is what we predicted that the low vowel is shorter than the high vowels with a falling tone.

The above durational and statistical results were obtained from the 0% to 100% timepoints, which is, more or less, under the consideration of exclusion of intrinsic, purely physiological, factors. In the case of IVD, however, the physiological and acoustic manifestations are our main concern. The whole length of T1 (from the 0% to 100+% timepoint), therefore, have been measured. The mean durations and SDs (n=6) in ms of T1 on [i] and [a] for individuals and the differences between them are given in Table 2, top half. It shows that for most speakers [i] is longer than [a]. In some cases the differences are large, 80 ms or more. Only two speakers, W2 and W5, have a longer [a] than [i]. The average durations and SDs of [i] and [a] for two sexes and for all speakers as a whole, their differences, and the 2-tail *t*-test results are given in Table 2, bottom half. It can be seen that for M, [i] is 45 ms longer than [a], and this is highly significant ($p < 0.001$). For W, [i] is longer than [a] by 23 ms, but the probability is marginal (*p* being slightly over 0.05 level). Overall, [i] is significantly longer than [a] by 35 ms ($p < 0.001$).

TABLE 2. Mean durations and SDs (in parentheses) in ms of T1 (0%-100+%) on [i] and [a]

	W2	W5	M2	M1	W3	M6	W1	M3	M4	W4	M5	*n*
[i]	*m*	289	287	310	310	237	370	402	303	377	367	369
	6											
	s	43	30	26	18	17	74	48	38	36	87	74
[a]	*m*	319	290	305	302	226	338	354	247	296	280	281
	6											
	s	38	13	13	18	21	25	72	54	40	44	30
[i]-[a]	-30	-3	5	8	11	32	48	55	80	87	88	

2-tail	[i]	[a]	*n*	[i]-[a]	*p*	DF
M	340 (56)	295 (41)	36	45	***	70
W	316 (77)	294 (59)	30	23	0.0556	58
ALL	329 (67)	294 (50)	66	35	***	130

Testing the second hypothesis. The durational data in ms for the rising tones (T2 and T3) are given in the top half of Table 3. The durational differences between high and low vowels can be read off in the '[i]-[a]' row. The probabilities from a series of 2-tail *t*-tests are given in the lower half of the table.

TABLE 3. Mean durations and SDs in ms of [i] and [a] with a rising tone T2 or T3

T2		W2	M4	W5	M3	M5	W3	M2	M1	M6	W4	W1
	n											
[i]	*m*	277	279	219	199	212	210	235	209	311	298	387
	6											
	s	29	30	21	9	17	8	29	17	38	53	46
[a]	*m*	343	332	264	219	230	223	245	198	288	275	348
	6											
	s	64	21	17	20	25	12	12	6	28	20	37
[i]-[a]		-66	-53	-45	-20	-18	-13	-10	11	23	23	39
T3		W2	M4	M3	W5	W3	W1	M5	W4	M2	M6	M1
[i]	*m*	245	295	220	221	230	368	233	282	283	315	205
	s	22	29	21	19	14	31	27	19	25	16	11
[a]	*m*	331	306	225	220	226	363	212	260	256	287	169
	s	28	25	21	19	17	20	20	28	17	29	16
[i]-[a]		-86	-11	-5	1	4	5	21	22	27	28	36

2-tail		[i]	[a]	*n*	[i]-[a]	*p*	DF
T2	M	241 (48)	252 (50)	36	-11	~	70
	W	278 (73)	291 (59)	30	-13	~	58
	All	258 (63)	270 (57)	66	-12	~	130
T3	M	258(47)	242(51)	36	16	~	70
	W	269(58)	280(62)	30	-11	~	58
	All	263(52)	259(59)	66	4	~	130

The duration of [i] is shorter that of [a] in T2 by 11 ms for M, by 13 ms for W, and by 12 ms for ALL. In T3, [i] is longer than [a] by 16 ms for M, but shorter than [a] by 11 ms for W; overall it is 4 ms longer. Two-tail *t*-tests did not show any significance between the durations of [i] and [a] in all the cases. This is what we predicted; low vowels may or may not longer than high vowels when bearing a rising tone.

CONCLUSIONS

The first two experimental hypotheses were confirmed, in as much as when the contextual influences were eliminated from speech production, the differences in vowel duration ascribable to physiological factors remained. The relation between vowel duration and tone shapes was found corroborated. In a falling tone, a high vowel is longer than a low vowel; in a rising tone, no significant differences exist between them.

REFERENCES

Abramson, Arthur S. 1960. *The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments*. PhD Dissertation, Columbia University.

Catford, J.C. 1977. *Fundamental problems in phonetics.* Edinburgh University Press.

Clark, J. & C. Yallop 1990. *An Introduction to Phonetics and Phonology.* Oxford: Basil Blackwell.

Condax, I. & R. Krones 1976. Duration of four vowels in manually produced synthetic speech. *Journal of Phonetics* 4:255-264.

Flanagan, J. 1972. *Speech Analysis, Synthesis and Perception.* Berlin: Springer-Verlag.

Hombert, Jean-Marie 1978. Consonant types, vowel quality, and tone. In Fromkin, V. (ed), *Tone: A Linguistic Survey.* New York: Academic Press. House, Arthur S. 1961. On vowel duration in English. *Journal of the Acoustical Society of Ameraica* 33:1174-8.

Keating, Patricia A. 1985. Universal phonetics and the organization of grammars. In V. Fromkin (ed) 1985. *Phonetic Linguistics.* Orlando: Academic Press.

Lehiste, Ilse 1970. *Suprasegmentals.* MIT Press.

Lindblom, B. 1967. Vowel duration and a model of lip-mandible coordination. *Speech Transmission Laboratory Quarterly Progress and Status Report* 4:1-29. Royal Institute of Technology, Stockholm.

Lieberman, P., Knudson, R. & J. Mead 1969. Determination of the rate of change of fundamental frequency with respect to subglottal air pressure during sustained phonation. *Journal of the Acoustical Society of America* 45: 1537-1543.

Peterson, G. & I. Lehiste 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32: 693-703.

Rose, P. 1981. *An Acoustically based Phonetic Description of the Syllable in the Zhenhai Dialect.* PhD thesis, The University of Cambridge.

Smith, B. 1987. Effects of bite block speech on intrinsic segment duration. *Phonetica* 44, 65-75.

Zhu, X. 1992. Intrinsic vowel F0 in a contour tone language. In R. Seidl (ed) *Proceedings of the Fourth Australian International Conference on Speech Science and Technology.* Melbourne.

Zhu, X. 1995. *Shanghai Tonetics.* PhD thesis, The Australian National University.