INTERACTION BETWEEN SUPRASEGMENTAL FEATURES

Jialu Zhang

Institute of Acoustics, Academia Sinica
Beijing, China

ABSTRACT

The interaction between the fundamental frequencies F0, the speech efforts and the speech rates were studied in different speech modes (loud, normal, soft, fast and slow). It is shown that 1) Both the F0 and the pitch range show a positive correlation with speech effort, but a rather very weak correlation in speech rate; 2) The onset F0 of sentences are mainly dependent on the speech efforts; 3) The declination rate of F0 is related to both speech effort and speech rate; 4) The lower boundary of pitch range moves up in fast speech and down in slow speech.

INTRODUCTION

The interaction between the voice source and the vocal tract has been studied over and over again since the beginning of this century. And some recent work [1] has laid the theoretical foundations for modeling of source-tract system. In order to get an insight into the model of speech production, we still need, however, to incorporate more theoretical knowledge and more experimental data in dynamic process of connected speech.

We thus have a general interest in the acoustic correlates of both segments [2] and suprasegmentals, because acoustic data of speech are not only easy to acquire but also necessary to deal with in speech signal processing systems.

This paper aims at examining the interrelation between the fundamental frequency F0, the speech effort and the speech rate, which are more important in suprasegmental level. And the declination of F0 in different speech modes were studied in this paper too.

METHODS

The speech material used in this study is four Chinese vowels /a,i,u and y/ with four tones —level, raising, dipping and falling. Each of the vowels with tone was embedded in a frame sentence "wǒ dú _ zì" (I utter the character _). Two speakers (one male and one female) who are native of Beijing uttered the testing material in five different speech modes and were recorded in a listening room in which the reverberation time is less than 0.5sec in the frequency band of speech. The speakers were asked to change their speech efforts in three levels—loud, normal and soft which correspond to 75, 65 and 55 db sound pressure level measured at 1 meter in front of speaker's lips respectively. The speakers were also asked to speed up and to slow down the speaking rate. The normal speech rate is about 4 syllables per second. As for the fast speech, the speakers uttered four sentences, in each sentence a testing vowel with tone was embedded, into one utterance rapidly, in order to check the relationship between the declination of F0 and the utterance length. The recorded material was digitized and analyzed using the API program of the ILS package.

RESULTS

1. The pitch range in different speech modes

The pitch range is determined between the highest and lowest fundamental frequencies of the utterance. The highest

F0max and the lowest F0min of the testing vowels with tones and of the whole frame sentences were measured individually. And the average pitch range of a vowel over four tones $\Delta \bar{F0}=\bar{F0}max-\bar{F0}min$ (the upper figures), and the relative pitch range F0max/F0min (the lower figures) in different speech modes are listed in Table 1. $\bar{F0}max$ and $\bar{F0}min$ stand for the average highest and lowest F0 values of a certain vowel over four different tones respectively. The pitch limit which was defined as the range from the ceiling to the floor of the voice which can be reached in different speech modes by the speakers is much larger than the average pitch range and not given in Table 1.

It was a pity that the API program did not work well and no data were given in Table 1,3,5 for female loud speech. From Table 1. it can be seen that 1) The pitch range is moved up and expanded as increasing the speech effort; 2) The pitch range is somewhat shrunken as

Table 1. The avreage pitch range of testing vowels in different speech modes in both $\bar{F0}max-\bar{F0}min$ (Hz) and F0max/F0min.

Vowel	Loud	Normal	Soft	Fast	Slow	Ave.
m	244-101	171-94	143-94	171-104	172-90	181-96
a	2.42	1.82	1.52	1.64	2.12	1.91
f		299-148	218-128	327-173	271-139	278-147
		2.02	1.70	1.89	1.95	1.89
m	280-105	199-102	143-97	183-113	193-95	199-103
i	2.67	1.95	1.45	1.62	2.03	1.93
f		330-150	222-128	329-204	287-146	293-157
		2.20	1.73	1.61	1.97	1.86
m	291-116	193-96	141-97	182-110	209-96	203-104
u	2.51	2.01	1.45	1.65	2.18	1.95
f		340-149	231-122	348-179	294-139	304-148
		2.28	1.29	1.94	2.12	2.05
m	288-122	194-100	141-95	190-113	196-93	202-105
y	2.36	1.94	1.48	1.68	2.11	1.92
f		315-150	234-126	362-187	294-148	301-154
		2.10	1.86	1.94	1.99	1.95
m	275-111	189-98	141-96	181-110	193-91	196-102
Ave.	2.49	1.93	1.48	1.65	2.11	1.92
f		321-149	226-126	341-186	286-143	294-152
		2.15	1.79	1.84	2.00	1.93

Note: m stands for male and f for female and Ave. for average.

Table 2. The average tone duration of vowels /a/ and /i/ v and the average frame sentence duriation s in different speech modes (ms).

Tone		Loud		Normal		Soft		Fast		Slow		Average	
		v	s	v	s	v	s	v	s	v	s	v	s
	m	272	1062	310	1111	298	1127	179	589	352	1530	282	1084
1	f	269	1168	250	1076	215	1053	144	599	365	1466	249	1072
	m	301	1059	298	1050	317	1095	196	615	448	1654	312	1095
2	f	314	1175	256	1114	253	1047	157	589	397	1536	275	1092
	m	282	1040	288	1082	295	1079	167	634	455	1596	297	1086
3	f	317	1187	262	1097	256	1040	125	525	362	1581	264	1086
	m	256	973	266	998	266	1031	154	627	375	1501	263	1026
4	f	263	1123	250	1066	205	1028	147	570	352	1456	243	1048

Note: 1 stands for level tone, 2 for raising, 3 for dipping and 4 for falling.

increasing speech rate; 3) The pitch limit of a speaker is about double what the speech range in normal voice; 4) The intrinsic pitch of vowels are well kept up in any speech mode; 5) The pitch range of male voice and of female voice are nearly the same in relative scale and semitone is a good measure of pitch for developing speech processing systems.

2. The duration of testing vowels and frame sentences in different speech modes

The durations of the testing vowels and the frame sentences were measured on speech waveforms. The results show that the duration ratio of testing vowel to frame sentence is nearly invariant in different speech modes. The average duration of low vowel /a/ and high vowel /i/ with the same tone type is called the tone duration and listed in Table 2.

It is worthy to be noted that the rank order of the tone durations from long to short is raising, dipping, level and falling and it is well kept up in any speech mode. In addition, the same relation exhibited in sentence level too. The fact that tone duration shows some intrinsic feature is identical with the results which were obtained from a statistical analysis of a large vocabulary[3].

3. The declination of F0 in different speech modes

The declination of F0 is easy to determine using the frame sentence designed in this paper, because of the sentence initial syllable with dipping tone and the final syllable with falling tone. Both dipping and falling tones can provide a local minimum in pitch contours. Then the baseline can be easy drawn by connecting the two minima. What is an exception to this is that when a vowel with dipping tone is embedded in the frame sentence, then the sentence final F0 may be higher than the local minimum of preceding dipping tone. That is due to the sharpening rule running on sentence intonation. In this case the baseline passes through the two minima of dipping tones.

Table 3. shows the F0 declination value △F0 and the sentence final fundamental frequency F0f in different speech modes. And the F0 declination value of the four sentence utterances and

Table 3. The F0 declination value △F0 and the sentence final fundemental frequency F0f for different vowels in different speech modes (Hz).

Vow.		Loud		Normal		Soft		Fast		Slow	
		△F0	F0f	△F0	F0f	△F0	F0f	△F0	F0f	△F0	F0f
a	m	51	103	19	95	14	88	26	96	27	82
	f			37	152	21	136	42	142	27	146
i	m	63	105	27	102	8	96	18	90	25	95
	f			49	159	25	147	27	146	23	145
u	m	51	116	23	96	7	98	28	92	23	95
	f			52	150	28	139	46	140	29	139
y	m	46	124	20	100	7	96	28	90	23	94
	f			42	151	27	150	59	144	15	149
Av.	m	53	112	22	98	13	95	26	92	25	92
	f			45	153	25	143	43	143	23	145

Note: Av. stands for average.

Table 4. The F0 declination value △F04 of the four sentence utterance and △F01 of the last sentence and the utterance final fundamental frenquency F0f for different vowels in fast speech (Hz).

Vowel	a		i		u		y		Ave.	
	m	f	m	f	m	f	m	f	m	f
△F04	40	64	40	61	58	100	51	81	44	77
△F01	26	42	18	27	28	46	28	59	25	44
F0f	96	142	90	146	92	140	90	144	92	143

of the last sentences and the utterance final F0 were shown in Table 4.

From Table 3. and Table 4. it can be seen that 1) The sentence final F0 value is invariant even for the four sentence utterance of fast speech; 2) The F0 declination values are speech mode dependent, the more effort has been made the more F0 declination value appears; 3) There is no distinct difference between male and female in relative value of F0 declination.

4. The onset F0 value of sentences in different speech modes

Table 5. shows the measurement results of the onset F0 of sentences in which different vowels were embedded in different speech modes.

Table 5. indicates that the onset F0 of sentences showed an obvious correlation with speech efforts but it is

Table 5. The onset F0 value of the sentences (Hz).

vowel	loud		Normal		Soft		Fast		Slow		Average	
	m	f	m	f	m	f	m	f	m	f	m	f
a	154		123	212	115	185	135	221	124	198	130	204
i	168		138	224	120	182	135	209	136	191	139	201
u	167		128	221	117	174	130	229	138	199	136	206
y	170		133	216	125	176	126	225	136	195	138	203
Ave.	165		131	218	119	179	132	221	134	196	136	204

independent of speech rates for both male and female. That means the onset F0 of sentences are mainly controlled by the subglottal pressure and laryngeal tension.

The average onset F0 value of sentence for a certain vowel over different speech modes were listed in last two columns in Table 5.

DISCUSSION

From Table 1. and 5. it is indicated that the speech efforts showed a good positive correlation with both the pitch ranges and the onset F0 of sentences, whereas speech rates did not exhibit a significant relation. The $\overline{F0}$max of the average pitch range for normal, fast and slow speech are nearly the same, because the speech level of them are equal. It appears that, speaker can manipulate the voice source and the vocal tract independently. The F0max of pitch range and the onset F0 of sentence are mainly controlled by subglottal pressure incorporated with laryngeal tension.

As for the F0 declination and the physiological process underlying it, the △F0 and F0f in different speech modes showed that the sentence final F0 is invariant but the sentence initial F0 is speech effort dependent and speech rate independent. So the declination rate is related to sentence length. In the case of four sentence utterance in fast speech, however, the F0 declined at different rates in different parts of the utterance. F0 declination value of the last sentence is above 50 percent of the total △F0 of the four sentence utterance. That means speaker can in some degree control the subglottal pressure to match the syntactic structures of the utterances. So that the F0 declination may be a passive phenomenon mixed with some active speaker control process.

It is worth notice that the intrinsic pitch of vowels and the intrinsic duration of tones are well kept up in different speech modes and showed somewhat effect on sentence level.

CONCLUSIONS

1. The fundamental frequency and the pitch range of speech show a strongly positive correlation with speech effort, but rather a very weak correlation in speech rate.
2. The onset F0 of sentences are mainly dependent on speech efforts and independent of speech rates.
3. The sentence final F0 is invariant in different speech modes, and the F0 declination rate is closely related to speech efforts and sentence lengths.

ACKNOWLEDGMENTS

The author would like to thank Prof. A. Fourcin for his encouragement and collaboration in study of voice characteristics. This work was supported by the National Project 863 of China.

REFERENCES

[1] Fant, G. (1982), "Preliminaries to analysis of the human voice source ", STL-QPSR 4/1982, 1-27.
[2] Zhang, Jialu (1987), "The intrinsic fundamental frequency of vowels and the effect of speech modes on formants", Proceedings of 11th ICPhS (Tallinn), Sy 3.5. 1-4.
[3] Zhang, Ning (1986), "On the relation between the duration and the phonetic structure of syllables", Master thesis, Institute of Acoustics, Academia Sinica.
[4] Carole, E. and et al. (1983), "Is declination actively Controlled?", in "Vocal fold physiology" edited by Titze, I.R. and Scherer, R.C., Colorado: The Denver Center for the Performing Arts, Inc.