

Specification of Chinese HV-Script Format

Ver.1.0.0

2004/02/13

YAMAHA Corporation

[Notes]

Operation is not guaranteed although this book is a document about HV-Script.

The contents of this document may be changed without notice.

Copyright of this document belongs to YAMAHA Corporation.

For reprinting or copying of this document, permission of Semiconductor Division of YAMAHA Corporation is needed.



Copyright (c)2004 YAMAHA CORPORATION

All rights reserved

Content

REVISION HISTORIY	3
1 INTRODUCTION.....	4
2 BASIC COMPOSITION	4
3 FORMAT	5
3.1 HEADER	5
3.2 BODY	5
3.2.1 Voice Character String.....	6
3.2.2 Rhythm Marks.....	9
3.2.3 Control Character String	12
3.2.4 Event	14
3.2.5 Comment Out.....	15
4 EXAMPLE OF HV-SCRIPT	16
5 CODE TABLE OF CHARACTER AND MARKS	17
6 DEFAULT VOICE MAP.....	23

Revision History

Ver.	Date	Contents
1.0.0.0	2004/02/13	First Edition

1 Introduction

HV-Script is the format for performing a voice synthesis which consists of the synthesis character string included a rhythm marks, a set up of pronounced voice, and a message to a playback application and etc. , and it serves as a text input in order to make the input easier by the users.

2 Basic Composition

HV-Script is composed of “Header” and “Body”.

Header is composed of the 4-byte of ASCII code characters, that displays a start of HV-Script and character code used by language and Body. Body is a character string written in the character code specified by Header, and consists of the vocal information and the playback information.

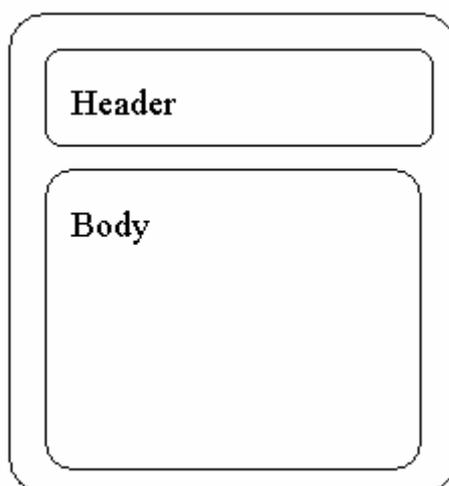


Fig. 1 Basic Composition of HV-Script

3 Format

3.1 Header

Header is expressed at the beginning of HV-Script. It is described with the 4-byte of alphabet capital letter (ASCII).

HV#C

HV# (0x48,0x56,0x23)	HV-Script start
C (0x43)	Indicate "Country", "Language" and "Character" Codes. In this case, it signifies the Chinese Language.

3.2 Body

Body is described by using character code of ASCII, and consists as follows.

- Voice character string : The character to utter is expressed.
- Rhythm mark : Marks which gives a clause(*1), asyllable(*2), and a rhythm (*3) to voice character string.
- Control character string : The volume, the pitch of sound, speaking speed, etc. are set up.
- Event : The message transmission to a playback application, and etc.
- Comment out : A comment can be written into HV-Script. The comment has no influence on the playback processing.

Character codes of GB2312, Big5, and ASCII is usable in Commnet out.

The alphabets used by Body are the small letter for Reading Marks, and capital letter for the other marks.

List of Marks and Character codes are shown in Table 9 to 15.

- *1 clause : The clause is a text unit which is divided from a head of clause till "Clause pause mark".
(For the details about "Clause pause marks", refer to the later in this document)

(Example) "HV#CW5ni2hao3-. W3huanving2. >2huanving2."
 clause clause clause

- *2 syllable : One "Reading mark" shows one syllable.
 *3 rhythm : The rhythm of the language made by the strength of sound, length, height, and etc.

3.2.1 Voice Character String

It is consisted of characters to utter and silence, and a character string which shows the punctuation of a clause like below.

- Reading Mark
- Long Vowel
- Silent
- Clause Punctuation Mark

3.2.1.1 Reading Mark [Alphabet String (PinYin)][Numeric [0-4](Voice Tone Mark)]

Input the syllable to pronounce by using PinYin. PinYin of inputtable syllable is shown in Table 9 to Table 12, and its ASCII code is shown in Table 13. Reading Mark consists of alphabet string, which expresses the syllable, and numeric value, which expresses the voice tone. Input alphabets with small letters. Input voice tone mark with numeric value for each syllable. Please input voice tone mark as the actual pronouncing voice tone.

(Example : 你好吗 : ni2hao3ma0)

Fig. 2 shows the pitch shifting of voice tone marks. Light Tone(轻声), which loses the voice tone, is described by 0, and its pitch shifts as shown in Fig. 3, according to the tone voice of the syllable right before. When a syllable except Third Tone exists just after Third Tone in one clause, only the first half of Third Tone shifts as shown in Fig. 2, and the second half becomes Half Third Tone with the lowest pitch. The pitch value with the voice tone mark shifts to the height of the syllable right before in a clause.

In addition, when Height Accent was added to Reading Mark or Long Vowel, voice tone mark does not shift the pitch of its Reading Mark utterance.

When extending the pronunciation by using Long Vowel, the pitch shifts as much as its length.

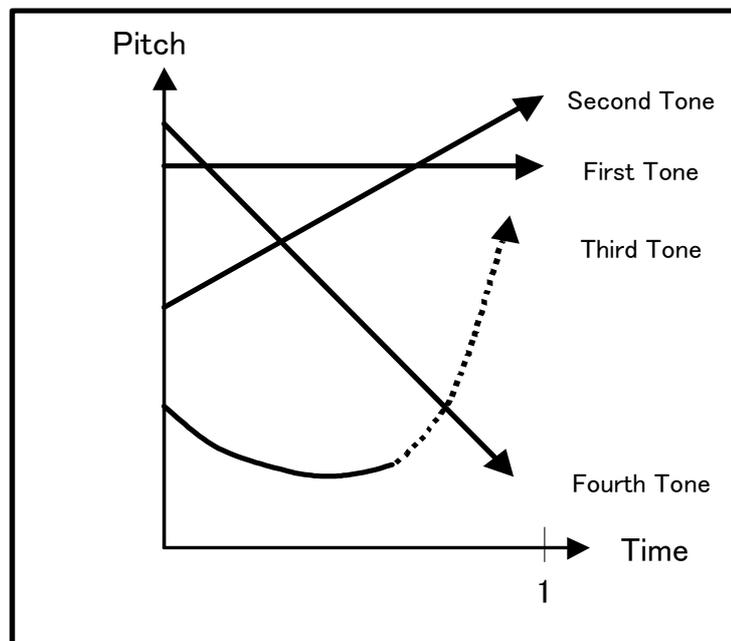


Fig. 2 Pitch Shifting of Voice Tone

“1” of horizontal axis shows utterance length of each character.
Utterance length of each character is different. (same in the followings)

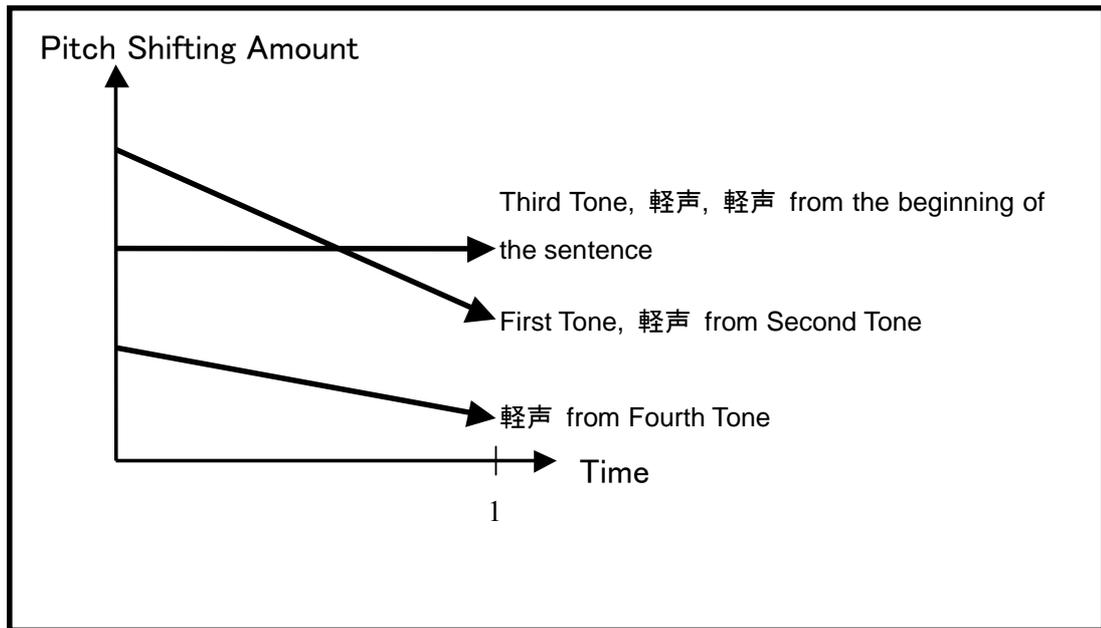


Fig. 3 Pitch shifting of 轻声 according to the Voice Tone of Syllable right before.

3.2.1.2 Long Vowel ["-","~"]

It pronounces the last “Reading mark” extended. It extends longer as it was inputted continuously.
“-” is the “long vowel” of fixed pitch, and “~” is the “long vowel” of vibrated pitch.

Example : “ni2----hao3~~~~”

3.2.1.3 Silent[" "]

A constant silent sound is inserted. The silent time becomes longer as the number inputted continuously.

3.2.1.4 Clause Punctuation Symbols [" , " , " . " , " ? " , " * "]

It shows the end of the clause.

“,” inputs one-length of “Silent”, and “.” inputs two-length of “Silent” compulsorily.

Symbol “?” raises the pitch and volume to the latest “Reading mark” or the latest “Long vowel”, and expresses interrogative rhythm. Pitch does not shift by voice tone mark of Reading Mark. The amount of change becomes large by attaching the numerical value of 1 to 99 back like “? 99”. Moreover, two-terms of “Silent” are inputted compulsorily. When “?” is inputted continuously like “?????”, each symbols are considered as the end of Clause. In this case, the amount of pitch and volume change is the same as one “?”, but Silent increases by the count of “?”.

A symbol “*” decreases a pitch and volume to a latest “Reading sign” or latest “Long vowel”, and expresses rhythm. Pitch does not shift by voice tone mark of Reading Mark. The amount of change becomes large by attaching the numerical value of 1 to 99 back like “* 99”. Moreover, two-terms of “Silent” are inputted compulsorily. When “*” is inputted continuously like “*****”, each symbols are considered as the end of Clause. In this case, the amount of pitch and volume change is the same as one “*”, but Silent increases by the count of “*”.

3.2.2 Rhythm Marks

Concerning to the utterance character string, it is expressed a rhythm by the change of pitch and volume.

3.2.2.1 Accent mark [" ' " , " ^ " , " / " , " _ " , " \$ " , " < " , " > " , " & " , " = "]

By describing it just before a "Reading mark" or a "Long vowel", it sets up a rhythm by the hight accents which change pitch (Table 1 / Fig. 4) , and by the strength accents which change volume (Table 2 / Fig. 5). In addition, it is accumulated the amount of change until a "Clause punctuation mark" or the symbols shown in Table 3 which returns a change.

Although it allows describing a "Height accent" and a "Strength accent" simultaneously to one of a "Reading mark" or a "Long vowel", when two or more different "Height accents" or "Strength accents" are described, it gives a priority to accents described later.

The combined use with the symbols, [/], [=], and [&], which return a change, is possible, and when symbols are inputted like ['/], it performs the change after returning the amount of shifts to 0.

Moreover, the amount of shifts becomes larger as the same mark was inputted continuously (Fig. 6), and it means the same thing as inputting the numerical value of 1 to 99 into just after the mark, such as [’99].

When "’" or "*" are effective to "Reading mark" or "Long vowel" which sets just after the "Accent mark", "Accent mark" will be skipped.

Table 1 Height accent mark

Symbol	Contents
'	Raises pitch at anlaut.
^	Raises pitch under pronunciation.
_	Decreases pitch at the beginning of a word.
\$	Decreases pitch under pronunciation.

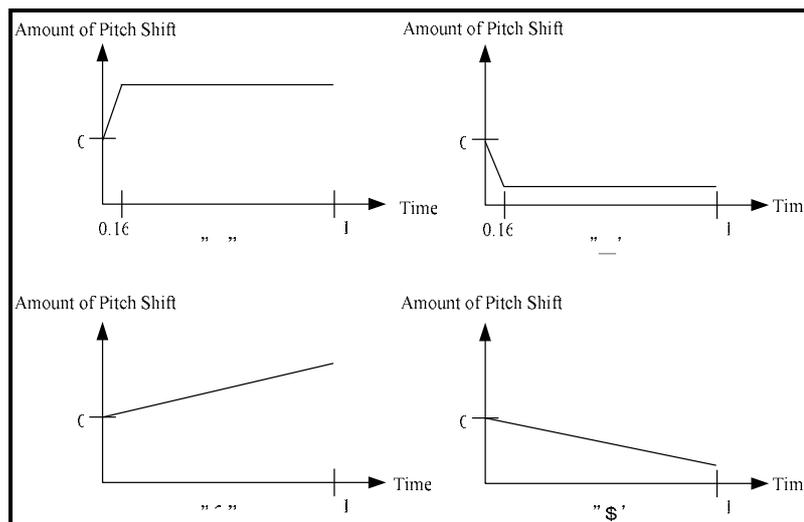


Fig. 4 Height Accent Symbols

Table 2 Strength accent mark

Symbol	Contents
<	Raises volume under pronunciation
>	Decreases volume under pronunciation.

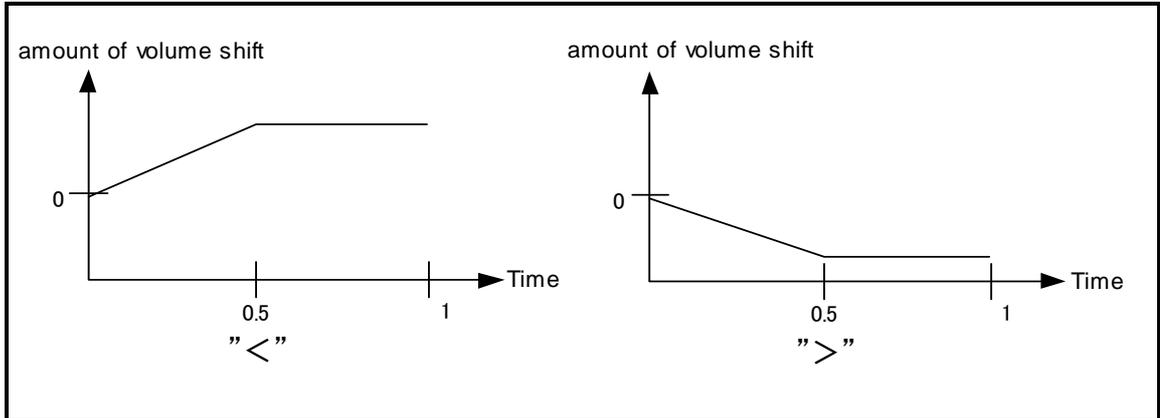


Fig. 5 Strength accent mark

Table 3 Clear accent shift

Symbol	Contents
/	Clears pitch shifted by "Accent mark".
=	Clears volume shifted by "Accent mark".
&	Clears pitch and volume shifted by "Accent mark".

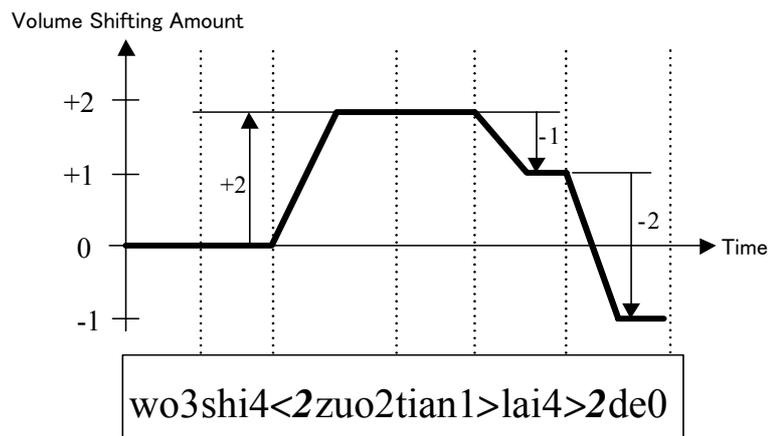


Fig. 6 Accent Symbols (example)

3.2.2.2 Rhythm of a clause ["@" , ":" , ";" , "!" , "+"]

It is Described at the head of a cause, and gives rhythm(tune) to the clause section uniformly. “Rhythm of clause” mark and its actions are shown in Table 4 / Fig. 7

It is effective only in the setup clause, and it has no change when it’s not described. Since “+” sets “Height accent mark” randomly to each “Reading mark” or “Long vowel”, “Height accent mark” written in the clause is ignored.

Table 4 Rhythm Symbols of a Clause

Symbol	Contents
@	It becomes higher with the 2nd "Read mark", and falls at the last.
!	It becomes lower with the 2nd "Read mark", and goes up by the last.
;	It becomes higher with the 2nd "Read mark."
:	It becomes lower with the 2nd "Read mark."
+	Random pitch

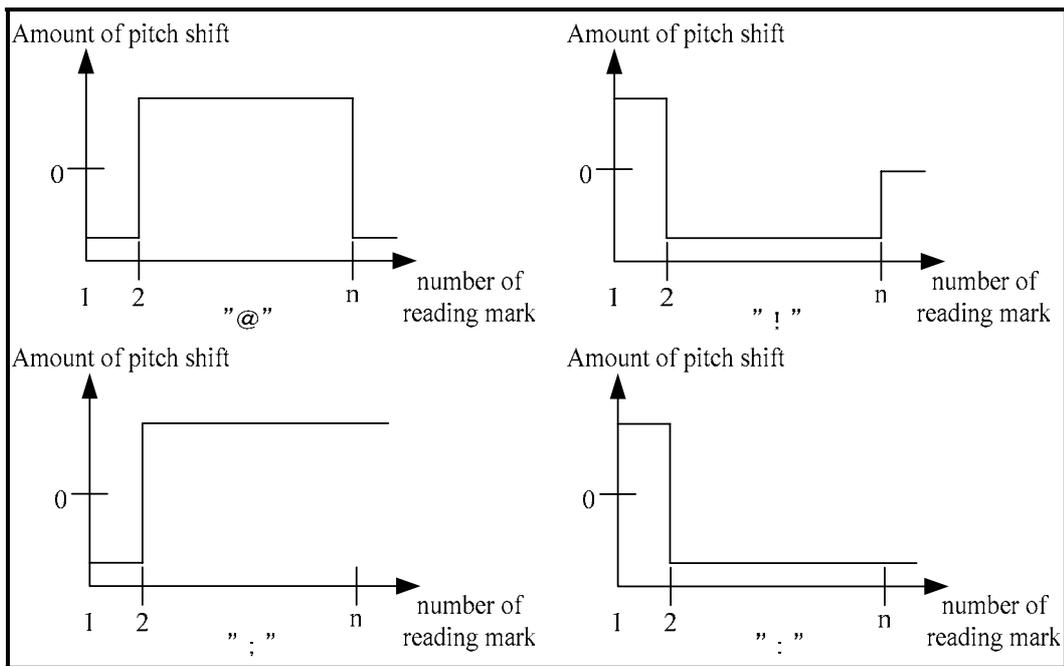


Fig. 7 Rhythm of a Clause (In case of a Reading mark number n)

3.2.3 Control Character String

By the use of the following marks, the change of a volume, pitch, and etc. are given. The same value is maintained until it changes the changed amount of change into the next time. Input Alphabets with Capital letters.

In addition, in the definition of the following setup values, it is indicated that the portion surrounded by "[]" is for the indispensable input, and the portion surrounded by "()" is for an arbitrary input.

3.2.3.1 Volume

V[value]

Numerical setting range : 1 to 5 (Default value : 4)

Explanation : Message to change volume

The amount of volume shift effected by each setting value is shown in Table 5.

It effects from the first "Reading mark" after the input location, until the location next Volume was set or the performance was finished

"Volume" between "Reading mark" and "Long vowel" or between "Long vowel" and "Long vowel" effects from the first "Reading mark" after "Long vowel".

Table 5 Table of Volume Change Parameter

Message	Amount of shift [dB]
V1	-24
V2	-18
V3	-12
V4	-6
V5	0

3.2.3.2 Pitch

[Music Scale] (Octave value)

Scale setting range : C, C#, D, D#, E, F, F#,G,G#,A,A#,B

Octave specification range : 1 to 3 (Default value : 2)

Explanation : Message to set height of voice sound to pronounce.

Specification of music scale and octave (Table 6). Height of octave is set automatically to the height which matches to the voice.

When an octave value is omitted, it would be set with the existing value.

When an pitch specification does not exist, it would be played with the default voice value.

It effects from the first "Reading mark" after the input location, until the location next "pitch" is specified, or the performance is finished.

"Pithe" between "Reading mark" and "Long vowel" or between "Long vowel" and "Long vowel" effects from the first "Reading mark" after "Long vowel".

Table 6 The amount of octave specification range shift

Octave specification range	Amount of shift
1	Default – 1 octave
2	Default
3	Default + 1 octave

3.2.3.3 Speed

S [Value]

Numerical specification range : 0 to 99 (Default value : 50)

Explanation : Message to shift utterance speed.

Utterance length becomes shorter of certain fixed quantity, as the specified value gains one. Also becomes longer as the value decreases.

It effects from the first “Reading mark”, “Long vowel”, “Silent”, “Clause punctuation mark”, and “Question mark” after input location, until the next “Speed” specified or the performance finished.

Speed like the following examples described in the Long-vowel sound becomes effective from the next “Reading Mark”.

(Example) HV#C S40ni2---S50--hao3ma0.

3.2.3.4 Utterance Length Unification

L [value]

Numerical setting range : 0 to 1 (Default value : 0)

Explanation : The switch to decide whether all utterance has same length or each one has peculiar length after the described position. ON/OFF of a switch is as in Table 7.

It effects from the first “Reading mark”, and “Long vowel” after input location, until the next “Utterance length unification” specified or the performance finished.

Utterance Length Unification like the following examples described in the Long-vowel sound becomes effective from the next “Reading Mark”.

(Example) HV#C L1ni2---L0--hao3ma0.

Table 7 Utterance Length Unification Setting Value

Numerical setting	Utterance length of a Reading mark
0	Peculiar length
1	Same length

3.2.4 Event

The message sent to a playback application and etc. at the time of playback of HV-Script.

3.2.4.1 Voice Quality Change

K[value], X[value]

Numerical setting range	:	0 to 15 (Default value : K0)
Explanation	:	It changes the voice quality of synthesized voice sound to the voice quality of numerical setting range. K0 to K15 are default voice. X0 to X15 are extended voice, and extended voice must be set up beforehand. It is effective from the first "Reading mark" and "Long Vowel" after inputted location, until the next "Voice quality change" specified or the performance finished. Refer to "Section 6 Default Voice Map" about contents of default voice

3.2.4.2 User Event

U[value]

Numerical setting range	:	0 to 9
Explanation	:	Message to execute the user specified event under the performance of HV-Script.

3.2.4.3 Height Accent / Degree of Rhythm Change of a Clause

W [value]

Numerical setting range	:	1 to 5 (Default value : 3)
Explanation	:	Message to change "Degree of rhythm change" of "Height accent mark" and "Rhythm mark of clause". It changes a lot as the specified value gains. Amount of shift for each numerical setting is as shown in Table 8. It is effective from the next "Reading mark" after the inputted location, until the next value specified or the performance finished.

Table 8 Height Accent / Rhythm Change Parameter of a Clause

Message	The degree of rhythm change (×2)
W1	1/6
W2	1/2
W3	1
W4	2
W5	5

3.2.5 Comment Out

It makes it possible to write comments into HV-Script.

[(character string)]

Explanation

: An area surrounded by "[]" is considered as a comment.
It is impossible to describe between nests or commands.
(Example: V[yin1 liang2]5)
CB2312, Big5, and ASCII are available.

4 Example of HV-Script

The following is the example of inputting HV-Script.

```
HV#CW5ni2hao3-.W3huan1ying2,>2huan1ying2,  
K1ni2hao3-.hao2jiu3S53bu2jian4le0.  
K0S50qing3zuo4.qing3he1<cha2.  
K1W5S40<2^3xie4&xie0.  
K0S50bie2ke4qi0.
```

5 Code Table of Character and Marks

Table 9 Syllable (1/4)

声母 韻母	無介音												
	a	o	e	-i[zh]	-i[z]	er	ai	ei	ao	ou	an	en	ang
b	ba*	bo					bai*	bei*	bao*		ban*	ben*	bang
p	pa*	po					pai*	pei*	pao*	pou	pan*	pen*	pang
m	ma*	mo	me				mai*	mei*	mao*	mou	man*	men*	mang
f	fa*	fo						fei*		fou	fan*	fen*	fang
d	da*		de*				dai*	dei*	dao*	dou*	dan*	den*	dang
t	ta*		te*				tai*		tao*	tou*	tan*		tang
n	na*		ne*				nai*	nei	nao*	nou	nan*	nen	nang
l	la*	lo	le*				lai*	lei	lao	lou	lan*		lang
g	ga*		ge*				gai*	gei	gao*	gou*	gan*	gen	gang
k	ka*		ke*				kai*	kei	kao*	kou*	kan*	ken	kang
h	ha*		he*				hai*	hei	hao*	hou*	han*	hen	hang
j													
q													
x													
zh	zha*		zhe*	zhi			zhai*	zhei	zhao*	zhou*	zhan*	zhen	zhang
ch	cha*		che*	chi*			chai*		chao*	chou*	chan*	chen*	chang
sh	sha*		she*	shi*			shai*	shei*	shao*	shou*	shan*	shen*	shang
r			re*	ri*					rao	rou*	ran	ren*	rang
z	za*		ze		zi*		zai*	zei*	zao*	zou	zan*	zen*	zang
c	ca*		ce		ci*		cai*		cao*	cou*	can*	cen*	cang
s	sa*		se		si*		sai*		sao	sou	san*	sen*	sang
zero	a*	o*	e			er	ai*	ei	ao	ou	an*	en	ang

儿(r)-ized input is possible for the syllable with '*'.

Table 10 Syllable (2/4)

声母 韻母	無介音		介音 i									
	eng	ong	i	ia	ie	iao	iou -iu	ian	in	iang	ing	iong
b	beng		bi*		bie	biao*		bian*	bin*		bing	
p	peng		pi*		pie	piao*		pian*	pin*		ping	
m	meng		mi*		mie	miao*	miu	mian*	min*		ming	
f	feng*											
d	deng	dong	di	dia*	die*	diao	diu	dian*			ding	
t	teng	tong	ti		tie*	tiao*		tian*			ting	
n	neng	nong	ni*		nie	niao*	niu*	nian*	nin*	niang	ning	
l	leng	long	li*	lia*	lie	liao*	liu*	lian*	lin*	liang	ling	
g	geng	gong										
k	keng	kong*										
h	heng	hong										
j			ji*	jia*	jie	jiao*	jiu*	jian*	jin*	jiang	jing	jiong
q			qi*	qia*	qie	qiao*	qiu*	qian*	qin*	qiang	qing	qiong
x			xi*	xia*	xie	xiao*	xiu	xian*	xin*	xiang	xing	xiong
zh	zheng	zhong*										
ch	cheng	chong										
sh	sheng											
r	reng	rong										
z	zeng	zong										
c	ceng	cong										
s	seng	song										
zero	eng		yi*	ya*	ye	yao	you	yan*	yin*	yang	ying*	yong*

儿(r)-ized input is possible for the syllable with '*'.

Table 11 Syllable (3/4)

声母 韻母	介音 u									介音 ü			
	u	ua	uo	uai	uei -ui	uan	uen -un	uang	ueng	ü	üe	üan	ün
b	bu												
p	pu												
m	mu*												
f	fu*												
d	du		duo		dui*	duan*	dun*						
t	tu		tuo*		tui*	tuan*	tun*						
n	nu		nuo			nuan				nv*	nve		
l	lu		luo			luan	lun*			lv*	lve		
g	gu	gua*	guo*	guai*	gui*	guan*	gun*	guang					
k	ku	kua*	kuo	kuai*	kui*	kuan*	kun*	kuang					
h	hu	hua*	huo*	huai*	hui*	huan*	hun*	huang					
j										ju*	jue	juan*	jun*
q										qu*	que	quan*	qun*
x										xu*	xue	xuan*	xun*
zh	zhu	zhua*	zhuo	zhuai*	zhui*	zhuan*	zhun*	zhuang					
ch	chu	chua	chuo*	chuai	chui	chuan	chun	chuang					
sh	shu	shua*	shuo*	shuai*	shui*	shuan*	shun*	shuang					
r	ru		ruo		ruì	ruan*	run						
z	zu		zuo*		zui*	zuan*	zun*						
c	cu		cuo*		cui	cuan*	cun						
s	su		suo*		sui	suan*	sun						
zero	wu	wa*	wo*	wai*	wei*	wan*	wen*	wang*	weng	yu*	yue	yuan*	yun*

儿(r)-ized input is possible for the syllable with '*.

Table 12 Syllable (4/4)
(Syllable used in spoken language)

yo
ng

Table 13 Alphabet

Alphabet	ASCII	Message	Alphabet	ASCII	Message
A	0x41	Pitch	g	0x67	PinYin
B	0x42	Pitch	h	0x68	PinYin
C	0x43	Pitch	i	0x69	PinYin
D	0x44	Pitch	j	0x6A	PinYin
E	0x45	Pitch	k	0x6B	PinYin
F	0x46	Pitch	l	0x6C	PinYin
G	0x47	Pitch	m	0x6D	PinYin
K	0x4B	Voice Quality change	n	0x6E	PinYin
L	0x4C	Utterance Length Unification	o	0x6F	PinYin
S	0x53	Speed	p	0x70	PinYin
U	0x55	User Event	q	0x71	PinYin
V	0x56	Volume	r	0x72	PinYin
W	0x57	Height Accent / Degree of Rhythm change of Clause	s	0x73	PinYin
X	0x58	Voice Quality change	t	0x74	PinYin
a	0x61	PinYin	u	0x75	PinYin
b	0x62	PinYin	v	0x76	PinYin
c	0x63	PinYin	w	0x77	PinYin
d	0x64	PinYin	x	0x78	PinYin
e	0x65	PinYin	y	0x79	PinYin
f	0x66	PinYin	z	0x7A	PinYin

Table 14 Symbols

Symbols	ASCII	Message
“(Space)	0x20	Silent
!	0x21	Rhythm of Clause
#	0x23	Pitch
\$	0x24	Height Accent
%	0x25	Rhythm of Clause
&	0x26	Clear shift of Height Accent and Strength Accent
'	0x27	Height Accent
*	0x2A	Clause Punctuation
+	0x2B	Rhythm of Clause
,	0x2C	Clause Punctuation
-	0x2D	Long Vowel
.	0x2E	Clause Punctuation
/	0x2F	Clear shift of Height Accent
:	0x3A	Rhythm of Clause
;	0x3B	Rhythm of Clause
<	0x3C	Strength Accent
=	0x3D	Clear shift of Strength Accent
>	0x3E	Strength Accent
?	0x3F	Clause Punctuation
@	0x40	Rhythm of Clause
[0x5B	Starts Comment Out
]	0x5D	Ends Comment Out
^	0x5E	Height Accent
_	0x5F	Height Accent
~	0x7E	Long Vowel

Table 15 Numeric Value

Numeric	ASCII	Message
0	0x30	Input numeric value
1	0x31	Input numeric value
2	0x32	Input numeric value
3	0x33	Input numeric value
4	0x34	Input numeric value
5	0x35	Input numeric value
6	0x36	Input numeric value
7	0x37	Input numeric value
8	0x38	Input numeric value
9	0x39	Input numeric value

6 Default Voice Map

Following table shows a default voice map.

Table 16 Default voice map

No.	Name	Basic Voice	Pitch Shift	Fixed Pitch	Prosodic Volume	Dimensions
0	Normal Man	m	0	—	E	Default Man
1	Normal Woman	w	0	—	E	Default Woman
2	Onih-san	m	0	—	E	Fine Young Man
3	Oneh-san	w	300	—	E	Gentle Young Lady
4	Boy	m	1200	—	E	Boy
5	Girl	w	1200	—	E	Girl
6	Radio Voice Man	m	100	—	E	Man voice from AM-radio
7	Radio Voice Woman	w	300	—	E	Woman voice from AM-radio
8	Hard-boiled	m	-1200	—	E	Man, Hard-boiled
9	Witch	m	0	—	E	Old Witch Woman
10	Hanazumari	m	0	—	E	Nasal Voice
11	Shitatarazu	m	200	—	E	Rat Voice
12	Water	m	0	—	E	Voice Of Underwater
13	Martian	m	0	—	E	Vibrated voice like an alien
14	Robot	m	0	80	E	Robot voice without intonation
15	Synth	m	1200	—	E	Synthesizer sound

* No 0 to 15 refer to K0 to K15 of Voice quality change (one of the event mark of HV-Script) .

* Basic Voice shows the sex of the base of the voice. [m] means a male, and [f] means a female.

* Pitch Shift shows the amount of pitch shift to the standard voice for each male and female. The unit is cent.

* Fixed Pitch is used to fix the pitch to play. The unit is Hz.

When the pitch is fixed, pitch shift set in HV-Script is ignored, and it is played with the designated pitch.

* Prosodic Volume is used to designate Enable / Unable of “Volume change specification”.

[E] means enable, and U means unable.

When the “Volume change specification” is designated to [Unable], “Strength accent” or “Volume specification” in HV-Script are ignored.