

日本語 HV-Script フォーマットガイド

第 1.0 版
2003 年 12 月 16 日

ヤマハ株式会社

<p>[注意事項]</p> <p>本書は HV-Script に関するドキュメントであり、再生システムの動作を保証するものではありません。</p> <p>本書の内容は予告なく変更される場合があります。</p>
<p>本書の著作権は、ヤマハ株式会社に帰属しています。</p> <p>本書の内容の転載・一部複製には、ヤマハ株式会社半導体事業部の許諾が必要です。</p>



Copyright (C) YAMAHA CORPORATION

目次

1	はじめに	4
2	基本構成	4
3	フォーマット	5
3.1	Header.....	5
3.2	Body.....	5
3.2.1	音声文字列.....	6
3.2.2	韻律記号	8
3.2.3	制御文字列.....	11
3.2.4	イベント	13
3.2.5	コメントアウト.....	14
4	HV-Script 例.....	15
5	文字・記号コード対応表.....	エラー! ブックマークが定義されていません。
6	デフォルト音色マップ.....	16

更新履歴

版	日付	内容
1.0	2003 年 12 月 16 日	初版

1 はじめに

HV-Script は、韻律記号を含んだ合成文字列、発音する音の設定、再生アプリケーション等へのメッセージから成る音声合成を行うためのフォーマットであり、ユーザによる入力を容易にするためテキスト入力となっている。

2 基本構成

HV-Script は Header と Body から構成される。

Header は HV-Script 開始と言語及び Body で使われる文字コードを表す 4Byte の ASCII 文字で構成される。

Body は Header で指定される文字コードで書かれた文字列で、発声情報や再生情報で構成される。

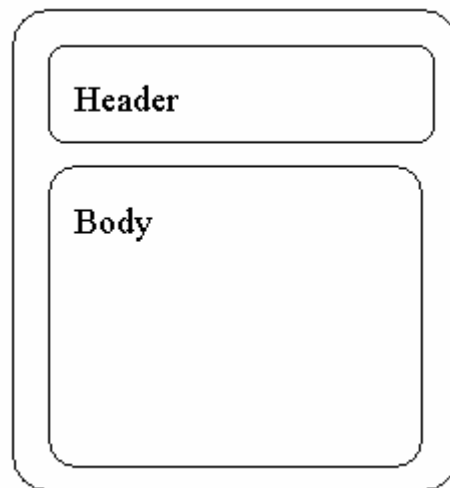


図 1 HV-Script 基本構成

3 フォーマット

3.1 Header

Header は 4Byte のアルファベット大文字(ASCII)で表される。

HV#J

HV# (0x48,0x56,0x23) HV-Script 開始

J (0x4A) 国・言語・文字コードを表す。この場合、日本語 Shift-JIS を意味する。

3.2 Body

Body は Header で指定された文字コード (Shift-JIS) を使って記述され、以下で構成される。

- ・音声文字列 : 発声する文字を表す。
- ・韻律記号 : 音声文字列に対し、文節(*1)単位または音節(*2)単位で韻律(*3)を与える記号
- ・制御文字列 : 音量や音の高さ、話速などを設定する。
- ・イベント : 再生アプリケーション等へのメッセージ送信。
- ・コメントアウト : HV-Script 中にコメントを書き込むことができる。コメントは再生処理に影響しない。

Body で使われるアルファベットは大文字小文字を同値に、またすべて全角・半角を同値として扱う。

記号・文字の Shift-JIS コード表を表 10～表 15 に示す。

*1 文節：文章の先頭から「文節の区切り記号」までで区切られる文章の単位のこと。

(「文節の区切り記号」については後述を参照。)

(例) HV#J こんにちわ。 げんき？ きょーは、 ひえるね。

文節 文節 文節 文節

*2 音節：1つの「読み記号」が1音節に当たる。

*3 韻律：音の強弱、長短、高低などによって作り出される言葉のリズム。

3.2.1 音声文字列

発声する文字及び無音、文節の区切りなどを示した文字列で、以下で構成される。

- ・読み記号
- ・長音
- ・無音
- ・文節の区切り記号

3.2.1.1 読み記号 [全角ひらがな(Shift-JIS)]

発声する文字。「読み記号」とその Shift-JIS コードを表 11～表 13 に示す。

発声する"音"の文字で入力すること。

例：「わたしは」 「わたしわ」

濁音は 1 文字で入力すること。

例：「し゛」 「じ゛」

但し、「う」は、濁点記号を用いて濁音化できる。濁点記号には「゚」(0x814A)を使用すること。

表 1 濁点の変換

入力	変換される読み記号
う゛ あ	ば
う゛ い	び
う゛ え	べ
う゛ お	ぼ
う゛ (上記以外の場合)	ぶ

3.2.1.2 長音 ["ー","～"]

直前の「読み記号」をのばして発声する。連続して入力した数だけ長くなる。

「ー」はピッチが一定の長音であるのに対し、「～」はピッチの揺れをもった長音になる。

例：「おはよー－－－－－」

3.2.1.3 無音 [" "]

一定の無音を挿入する。連続して入力した数だけ無音時間が長くなる。

3.2.1.4 文節の区切り記号 [", ", ", ", ", ". ", ". ", "? ", "* "]

文節の終わりを示す。

"、"、"、"では「無音」1つ、"。"、"。"では2つ分の長さの「無音」が強制的に挿入される。

"?"は直前の「読み記号」または「長音」に対してピッチと音量を上げ、疑問の韻律を表現する。「? 9 9」のように後ろに1から99の数値を付けることでピッチ及び音量変化量が大きくなる。また2つ分の長さの「無音」が強制的に挿入される。「?????」のように"?"を連続して入力した場合、それぞれが文節の終わりを示しているともみなす。この場合、ピッチ及び音量変化量は"? 1つと変わらないが、無音は"? の数だけ増加する。

"*"は直前の「読み記号」または「長音」に対してピッチと音量を下げて韻律を表現する。「* 9 9」のように後ろに1から99の数値を付けることでピッチ及び音量変化量が大きくなる。また2つ分の長さの「無音」が強制的に挿入される。「*****」のように"*"を連続して入力した場合、それぞれが文節の終わりを示しているともみなす。この場合、ピッチ及び音量変化量は"* 1つと変わらないが、無音は"* の数だけ増加する。

3.2.2 韻律記号

音声文字列に対し、文節単位または音節単位で韻律を与える。

3.2.2.1 アクセント記号 ["'","^","/","_","\$","<",">","&","="]

「読み記号」または「長音」の直前に記述し、ピッチを変化させる高低アクセント(表 2,図 2)、音量を変化させる強弱アクセント(表 3,図 3)で韻律の設定を行う。「文節の区切り記号」または表 4 に示す変化を戻す記号までは変化量を累積していく。

1つの「読み記号」または「長音」に対し、「高低アクセント記号」と「強弱アクセント記号」を同時に記述することを許すが、複数の異なる「高低アクセント記号」または「強弱アクセント記号」を記述した場合は後着優先となる。ただし、「/」「=」「&」の変化を戻す記号との併用は可能であり、「/」のように入力された場合は、変化量を0に戻してから変化を行う。

また、連続して同じ記号を入力すると変化量が大きくなり(図 4)、「99」のように後ろに1から99の数値をつけることにより連続入力と同じ意味を持つ。

「アクセント記号」の直後の「読み記号」または「長音」に対して"?"または"*"が有効な場合、「アクセント記号」は読み飛ばしされる。

表 2 高低アクセント記号

記号	内容
'	語頭でピッチを上げる
^	発音中ピッチを上げる
_	語頭でピッチを下げる
\$	発音中ピッチを下げる

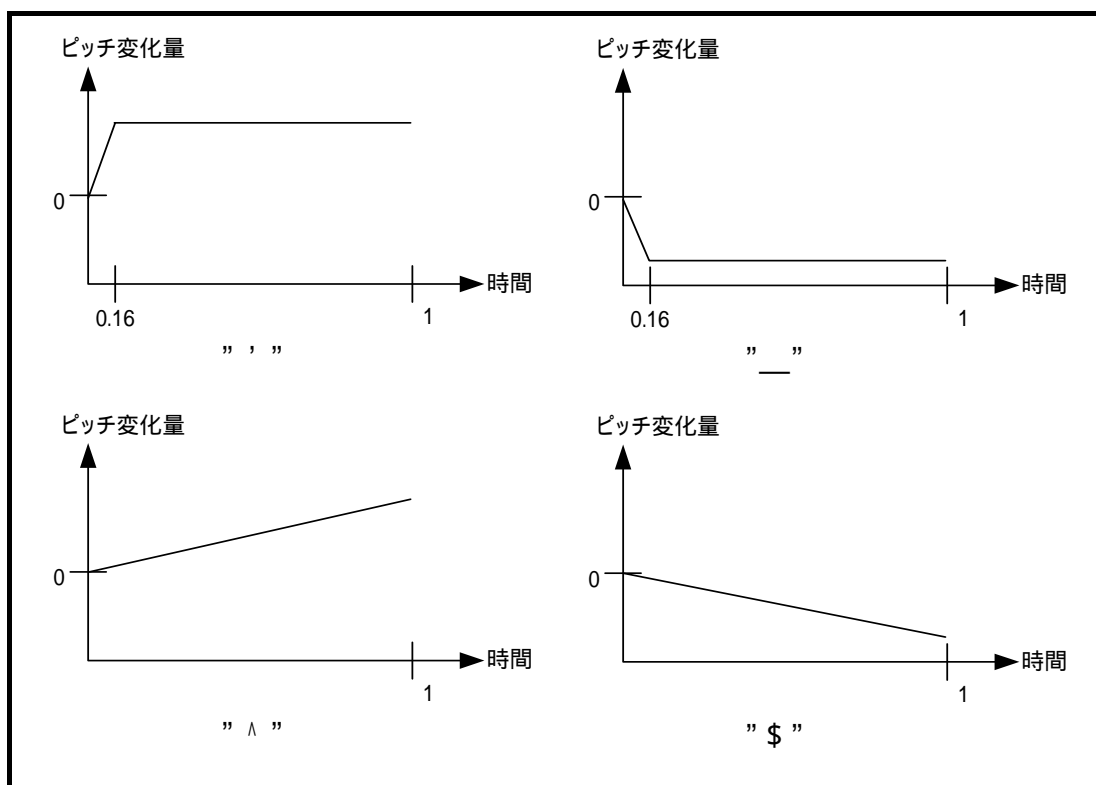


図 2 高低アクセント記号

横軸の"1"は文字ごとの発話長を表す。発話長は文字ごとに異なる。

表 3 強弱アクセント記号

記号	内容
<	発音中に音量を上げる
>	発音中に音量を下げる

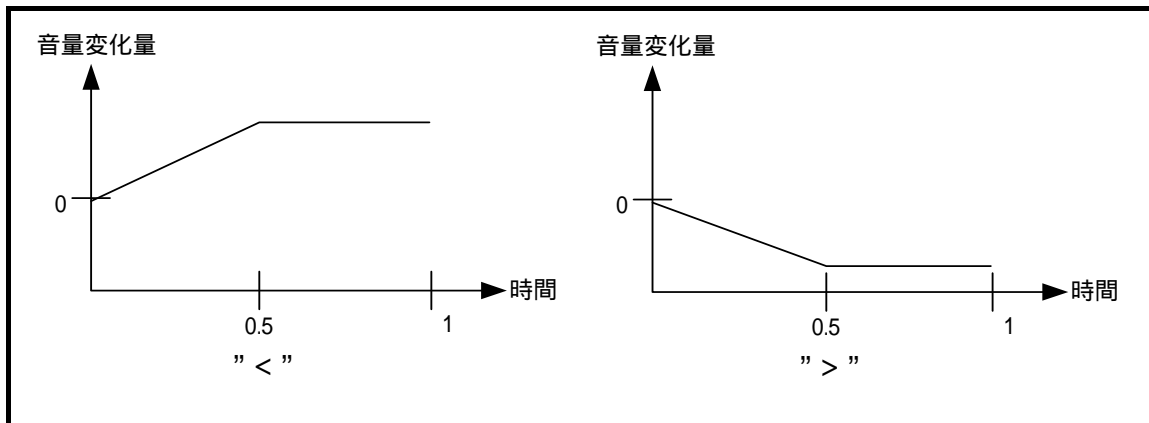


図 3 強弱アクセント記号

横軸の"1"は文字ごとの発話長を表す。発話長は文字ごとに異なる。

表 4 アクセントシフトクリア

記号	内容
/	アクセント記号で変化させたピッチをクリアする。
=	アクセント記号で変化させた音量をクリアする
&	アクセント記号で変化させたピッチと音量をクリアする。

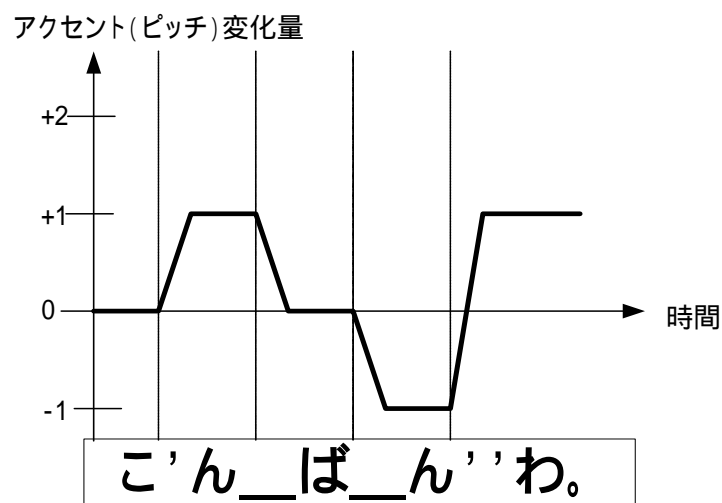


図 4 アクセント記号(例)

3.2.2.2 文節の韻律 ["@", ":", ";", "!", "+"]

文節の先頭に記述し、文節区間に一様にかかる韻律(調子)を与える。「文節の韻律記号」と作用を表 5, 図 5 に示す。

設定した文節にのみ有効であり、記述しない場合は変化なし。

“+”は文節の各「読み記号」、「長音」に対してランダムに「高低アクセント記号」を設定するため、文節に書かれた「高低アクセント記号」は無視される。

表 5 文節の韻律記号

記号	内容
@	2 つ目の読み記号で高くなり最後で下がる
!	2 つ目の読み記号で低くなり最後で上がる
;	2 つ目の読み記号で高くなる
:	2 つ目の読み記号で低くなる
+	ランダムピッチ

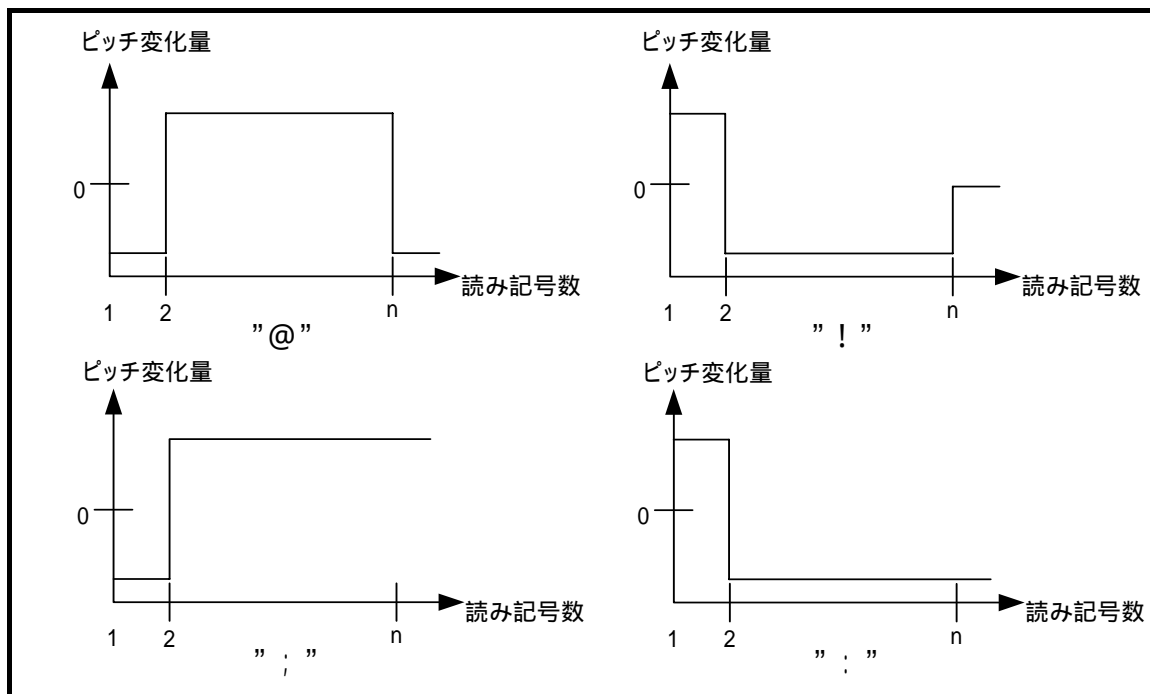


図 5 文節の韻律(読み記号数 n の場合)

横軸の"1"は文字ごとの発話長を表す。発話長は文字ごとに異なる。

3.2.3 制御文字列

以下の記号を用いて、音量やピッチなどの変化を与える。変化させた変化量は次に変更するまで同じ値が維持される。

なお、以下の設定値の定義で"[]"で囲まれている部分は必須入力、"()"で囲まれている部分は任意入力であることを示す。

3.2.3.1 音量

V[数値]

数値設定範囲	: 1-5 (初期値: 4)
説明	: 音量を変えるメッセージ。 各設定数値に対する音量変化量は表 6 のとおり。 入力位置の次の「読み記号」から、次に「音量」が指定されるか再生が終了するまで有効。 「読み記号」と「長音」、「長音」と「長音」の間の「音量」は、「長音」以降の次の「読み記号」から反映される。

表 6 音量変化パラメータ表

メッセージ	変化量 (dB)
V1	-24
V2	-18
V3	-12
V4	-6
V5	0

3.2.3.2 ピッチ

[音階] (オクターブ数値)

音階設定範囲	: C, C#, D, D#, E, F, F#, G, G#, A, A#, B
オクターブ指定範囲	: 1-3 (初期値: 2)
説明	: 発音する声の高さを設定するメッセージ。 音階とオクターブ (表 7) の指定。オクターブの高さは、音色に合った高さに自動的に設定される。 オクターブ数値を省略した場合はそれまでの値で設定される。 ピッチの指定がない場合は、音色のデフォルト値で再生される。 入力位置の次の「読み記号」から、次に「ピッチ」が指定されるか再生が終了するまで有効。 「読み記号」と「長音」、「長音」と「長音」の間の「ピッチ」は、「長音」以降の次の「読み記号」から反映される。

表 7 オクターブ指定範囲変化量

オクターブ指定範囲	変化量
1	デフォルト-1 オクターブ
2	デフォルト
3	デフォルト+1 オクターブ

3.2.3.3 速度

S [数値]

- 数値指定範囲 : 0-99 (初期値 : 50)
- 説明 : 発話速度を変えるメッセージ。
 指定数値が 1 増えるごとに発話長が一定量短くなり、減るごとに長くなる。
 入力位置の次の「読み記号」、「無音」、「区切り記号」から、次に「速度」が
 指定されるか再生が終了するまで有効。
 以下の例のような、長音中に記述した速度についての動作は保証しない。
 (例) HV#J あ S40ーーーS50ー。

3.2.3.4 発話長統一

L [数値]

- 数値指定範囲 : 0-1 (初期値 : 0)
- 説明 : 記述された位置以降全ての発話長を同じにするか、それぞれ固有の長さにするかのスイッチ。スイッチの ON/OFF は表 8 の通り。
 入力位置の次の「読み記号」、「無音」から、次に「発話長統一」が指定されるか再生が終了するまで有効。
 以下の例のような、長音中に記述した発話長統一についての動作は保証しない。
 (例) HV#J あ L1ーーーL0ー。

表 8 発話長統一設定値

数値指定	読み記号の発話長
0	固有の長さ
1	同じ長さ

3.2.4 イベント

HV-Script の再生時に再生アプリケーション等にするメッセージ。

3.2.4.1 声質変更

K[数値]、X[数値]

数値設定範囲	: 0-15 (初期値: K0)
説明	: 合成音声の声質を数値設定の声質に変える。 K0-K15 はデフォルト音色。X0-X15 は拡張音色であり、あらかじめ拡張音色が設定されていないといけない。 入力位置の次の「読み記号」, 「長音」から、次に「声質変更」が指定されるか再生が終了するまで有効。 デフォルト音色の内容については、6章「デフォルト音色マップ」を参照。

3.2.4.2 高低アクセント・文節の韻律変化度

W[数値]

数値設定範囲	: 1-5 (初期値: 3)
説明	: 「高低アクセント記号」, 「文節の韻律記号」の韻律変化度を変えるメッセージ。 指定数値が大きいほど大きく変化する。各設定数値における変化量は表 9 のとおり。 入力位置の次の「読み記号」, から、次に値が指定されるか再生が終了するまで有効。 以下の例のような、長音中に記述した高低アクセント・文節の韻律変過度についての動作は保証しない。 (例) HV#J あ'W5ーーーW3ー。

表 9 高低アクセント・文節の韻律変化パラメータ表

メッセージ	変化量(倍)
W1	1/6
W2	1/2
W3	1
W4	2
W5	5

3.2.5 コメントアウト

HV-Script 中にコメントを書き込むことを可能にする

[(文字列)]

説明

: "[]" で囲まれた領域はコメントとして扱う。

入れ子や各命令の間に記述する(例: V[音量]5)ことは不可。

4 HV-Script 例

HV-Script の入力例を以下に示す。

HV#J@S54 おはよ^~ <、ございます U1。 お'<げ_ん=きですか？ S56 きょ_おわ[「きょうは」は「きょおわ」と入力する。]V5s51 とっっっ'ても,S 54'い/いて\$ん_き、です_ね。L1W5K4 そーですネ 。

5 デフォルト音色マップ

以下に、デフォルト音色の特徴について記載する。

表 10 デフォルト音色マップ

No.	Name	Dict	Pitch Shift	Fixed Pitch	Prosodic Volume	特徴
0	Normal Man	m	0	-	E	標準男性の声
1	Normal Woman	w	0	-	E	標準女性の声
2	Onih-san	m	0	-	E	元気な男性の声
3	Oneh-san	w	300	-	E	やさしい女性の声
4	Boy	w	700	-	E	男の子の声
5	Girl	w	1200	-	E	女の子の声
6	Radio Voice Man	m	100	-	E	AM ラジオ風男性の声
7	Radio Voice Woman	w	300	-	E	AM ラジオ風女性の声
8	Hard-boiled	m	-1200	-	E	低く渋い男性の声
9	Witch	m	0	-	E	魔法使いの老婆のような声
10	Hanazumari	m	0	-	E	鼻づまりな男性の声
11	Shitatarazu	m	200	-	E	舌足らずな男性の声
12	Water	m	0	-	E	水中にいるような声
13	Martian	m	0	-	E	宇宙人のような震えた声
14	Robot	m	0	80	E	抑揚のないロボットボイス
15	Synth	m	1200	-	E	シンセ音

No の 0 ~ 15 は、声質変更（HV-Script のイベント記号の 1 つ）の K0 ~ K15 に対応する。

Dict は、音色の基本となる性別で、m：男性 w：女性 を意味する。

Pitch Shift は、男女それぞれ、標準の声に対するピッチの変化量を表す。単位は cent。

Fixed Pitch は、再生するピッチを固定する場合に使用する。単位は Hz。

ピッチを固定した音色の場合、HV-Script 中のピッチ変化は無視され、指定されたピッチで再生される。

Prosodic Volume は、HV-Script 中の音量変化指定の有効 / 無効に使用し、

E：Enable（有効） U：Unable（無効）を意味する。

音量変化指定が無効の音色の場合、HV-Script 中の強弱アクセントや音量指定は無視される。