音声言語情報処理 45－11
（2003．2．7）

音声認識読み記号および音声関連ソフトウエアに係わる用語の試行標準案

松浦 博[1] 西本 卓也[2] 金子 宏 磯谷 亮輔[3]
石川 泰[4] 西村 雅史[5] 伊藤 克亘[6] 新田 恒雄[7]
1 東芝研究開発センター 2 東京大学大学院 情報理工学系研究科 3 NECマルチメディア研究所
4 三菱電機情報技術総合研究所 5 日本IBM東京基礎研究所 6 産業技術総合研究所 7 豊橋技術科学大学大学院工学研究科
〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘1-1 E-mail : nitta@tutkie.tut.ac.jp[7]

あらまし： 本報告では，平成14年に発足した情報処理学会試行標準専門委員会下のWG4小委員会（音声言語処理インタフェース）が策定した，(A) 音声認識読み記号（エンドユーザ向け）の試行標準案，および (B) 音声認識・合成に係わる用語（開発者向け）の試行標準検討案の内容を紹介する。

キーワード： 音声言語処理，標準化，用語，読み記号，音声認識，音声合成

IPSJ Trial Standard Concerning Spoken Language Interface

Hiroshi MATSU'URA[1], Takuya NISHIMOTO[2], Hiroshi KANEKO, Ryosuke ISOTANI[3],
Yasushi ISHIKAWA[4], Masafumi NISHIMURA[5], Katunobu ITOU[6], and Tsuneo NITTA[7]
1 Toshiba Corp., 2 Univ. of Tokyo, 3 NEC Corp.,
4 Mitsubishi Electric Corp., 5 IBM Japan Ltd., 6 AIST, 7 Toyohashi Univ. of Tech.
1-1 Hibariga-oka, Tempaku, Toyohashi, 441-8580 JAPAN E-mail : nitta@tutkie.tut.ac.jp[7]

Abstract: This paper describes the activity of IPSJ Trial Standard WG4 that has the objectives of standardization for spoken language interface. In this report, we introduce two working drafts: (A) a phonemic system described in Kana used by end-users of ASR, (B) Terminology on ASR and TTS used by application developers.

Key words: Spoken Language Processing, Standardization, Terminology, Phonemic System, Speech Recognition, Speech Synthesis

1．はじめに

平成14年より情報処理学会に学会試行標準専門委員会（委員長石崎俊慶応義塾大学教授）が発足し，この委員会の下でWG4小委員会：音声言語処理インタフェースが発足し活動している[1]。学会試行標準は，情報処理学会のWebで公開され，国内外の規格化作業に役立てて頂くことを目指している。WG4小委員会では，今回，(1) 音声認識読み記号の試行標準案，(2) 音声認識・合成に係わる用語の試行標準検討案を策定した。本報告は，音声言語処理研究・開発に携わる多くの方々の意見を収集することを目的にまとめたものである。

2．音声認識読み記号

音声認識の対象単語に用いられる読み記号は，表記の規準がないため，製品での対応は各社まちまちである。また，音

声認識のAPIにおいても，読みまでには踏み込まず，使用記号を規定するにとどめているケースが見受けられる。読みの付与が使用者にゆだねられると混乱を招く恐れが生じる。こうした事情から，WG4 は読み記号を早期に情報処理学会試行標準とする作業を開始した。今後，業界団体等において規格化へ向けた議論が高まることを望むと同時に，音声認識用API などへも普及することを期待したい。

2.1 標準化の範囲

● 「音声認識読み記号」として表1に示す各記号を標準とする。網掛けされたアルファベット記号は発音の参考であり，標準化の対象ではない。標準に準拠した認識システムは，この範囲で入力された記号を少なくとも受理するか，あるいは受理できない旨のエラーを返すものとする。ただし，音声認識システムによっては受理した結果，必ずしも異なる記号が異なる発音にならないことがあっても，これは標準の範囲内として許容する。

● 音声合成を用いて認識結果をトークバックする（音声で知らせる）場合などを考慮し，JEIDA 規格である「日本語テキスト音声合成用記号の規格」(JEIDA-62-2000)との連携を図る。こうした利用と，開発者がより詳細に読みを記述することを目的に表2の「音声認識詳細読み記号」を設ける。

● 音声認識読み記号の並び方に関する規則や制約を細かく規定すると，ルールが複雑になりすぎ，かえって利用することが困難になる。そのため，音声認識読み記号の並び方に関する規定は特に設けない。すなわち，表1に示す「音声認識読み記号」の並べ方および表2に示す「音声認識詳細読み記号」の並べ方は，各々の範囲内で自由に設定できるものとする。また，音声認識システムによっては，並べ方に一部制限を設けることも考えられるが，標準の範囲内として許容する。

2.2 音声認識読み記号に関する説明

エンドユーザが使用できることを念頭においている。

<規　定>

● 「ひらがな」をベースに，表 1 に示す「音声認識読み記号」を定める。

● 鼻濁音化の記号（例えば「゜」）については，エンドユーザは鼻音と鼻濁音を必ずしも区別できないと考えられるため，採用しない。

● 無声化を表す記号も同様の理由で採用しない。

● 表1のアルファベットは発音の参考である。ただし，現代かなづかい[2] に従うと読み記号は発音と一致しないことがある。このような場合は，例えば現代かなづかいの本則で「とうきょう」「ていせい」などと記述されるとき，音声認識システム側で長音化処理を行い，それぞれ「とーきょー」「てーせー」等の発音がありうると解釈し処理するものとする。

● 一方，「小売」→「こうり」のように長音化しない場合がある。このような場合のため，長音化しないことを示す区切り記号として「・」（中黒あるいは中点）の使用をオプションとして認める。すなわち，「小売」は「こ・うり」と記述できる。なお，音声認識システムによっては「・」の代わりに「／」ほかの採用を宣言して用いても良い。

● 現代かなづかいの例外と位置付けられるもののうち，

表1　音声認識読み記号

あ a	い i (ぁ)	う u (ぅ)	え e	お o	や ya	ゆ yu	いぇ ye	よ yo	わ wa	うぃ wi	うぇ we	うぉ wo
か ka	き ki	く ku	け ke	こ ko	きゃ kya	きゅ kyu		きょ kyo	くぁ kwa	くぃ kwi	くぇ kwe	くぉ kwo
さ sa	し shi	す su	せ se	そ so	しゃ sha	しゅ shu	しぇ she	しょ sho				
た ta	ち chi	つ tsu	て te	と to	ちゃ cha	ちゅ chu	ちぇ che	ちょ cho	つぁ tsa	つぃ tsi	つぇ tse	つぉ tso
な na	に ni	ぬ nu	ね ne	の no	にゃ nya	にゅ nyu		にょ nyo				
は ha	ひ hi	ふ hu	へ he	ほ ho	ひゃ hya	ひゅ hyu		ひょ hyo	ふぁ fa	ふぃ fi	ふぇ fe	ふぉ fo
ま ma	み mi	む mu	め me	も mo	みゃ mya	みゅ myu		みょ myo				
ら ra	り ri	る ru	れ re	ろ ro	りゃ rya	りゅ ryu		りょ ryo				
が ga (nga)	ぎ gi (ngi)	ぐ gu (ngu)	げ ge (nge)	ご go (ngo)	ぎゃ gya (ngya)	ぎゅ gyu (ngyu)		ぎょ gyo (ngyo)	ぐぁ gwa (ngwa)			
ざ za	じ ji	ず zu	ぜ ze	ぞ zo	じゃ ja	じゅ ju	じぇ je	じょ jo				
だ da	ぢ di	づ du	で de	ど do		でゅ dyu						
ば ba	び bi	ぶ bu	べ be	ぼ bo	びゃ bya	びゅ byu		びょ byo				
ぱ pa	ぴ pi	ぷ pu	ぺ pe	ぽ po	ぴゃ pya	ぴゅ pyu		ぴょ pyo				
	てぃ ti	とぅ tu				てゅ tyu						
	すぃ si					ふゅ fyu						
	ずぃ zi				ぢゃ dya	ぢゅ dyu	ぢぇ dye	ぢょ dyo				
ん n	っ (促音)	ー (長音)		・ (区切り)		ぢ di	づ du	を o				

促音：　っ(JIS 0x2443)
区切り：・(JIS 0x2126)
長音：　ー(JIS 0x213c)
[注1]　表記例のコードは全角平仮名(JIS X 0208-1983「情報交換用漢字符号系」)を使用するが，実際に使用するコード系まで規定するものではない。
[注2]　この表に定義されていない「ゐ」「ゑ」等の仮名記号は使用しない。
[注3]　長音記号として「－（マイナス　JIS 0x215d)」は使用しない。
[注4]　網掛けされたアルファベット表記は発音の参考であり，音声認識読み記号の規定対象ではない。

「wa」と発音する助詞「は」，「e」と発音する助詞「へ」は発音どおりにそれぞれ「わ」，「え」と記述することとする。したがって，この記述だけは現代かなづかいの規定と一致しないことになる。

● 他の現代かなづかいについてはこれに従うこととする。この規定は認識システムからみても問題は少ない。例えば，助詞「を」は一般に「o」，場合によって「wo」に近く発声したとしても，ユーザは「を」と記述すれば良い。

＜補足説明＞

● エンドユーザが利用できる簡易な形式を目指し，より馴染みやすい「ひらがな」を用いた。

● JEIDA 規格 日本語テキスト音声合成用記号の規格(JEIDA-62-2000) に掲載された表 2-1 をベースとして作成している。また，国語審議会の「外来語の表記」[3] を参考に，一般的な外来語表記をカバーするようにした。

● 音声認識読み記号から発音への変換精度を上げる様々な工夫（音声合成を用いた読み上げによる確認，発声登録，単語辞書との突合せなど）は，音声認識システムの製作者にゆだねた。入力誤り，ユーザごとの発音の揺れなどへの対処も，音声認識システムの製作者にゆだねた。

● 理解の手助けとして以下に例を示す。ここで（ ）に記した発音表記は，認識システムが読み記号を解釈し内部記号に置き換えた一例である。

（凡例）画面に表示する文字　音声認識読み記号の表記（ある認識システムで解釈された発音表記の例）

東京　とうきょう (to-kyo-,toukyou)

武蔵新城　むさししんじょう
(musashishinnjo-, musashishinnjou)

丸の内　まるの・うち (marunouchi)

丸の内　まるのうち (maruno-chi,marunouchi)

大阪阿倍野橋　おおさか・あべのばし (o-sakaabenobashi)

経営　けいえい (ke-e-,keiei)

経営　け・いえい (keie-,keiei)

ユーザ　ゆーざー (yu-za-) …… 外来語もひらがなで記述

秋田　あきた (ak(i)ta) …… 記号列から音声認識システムが無声化を推測した例

鼻血　はなぢ (hanaji)

続く　つづく (tsuzuku)

私は　わたしわ (watashiwa)

表２　音声認識詳細読み記号

ア	イ	ウ	エ	オ	ヤ	ユ	イェ	ヨ	ワ	ウィ	ウェ	ウォ
a	i	u	e	o	ya	yu	ye	yo	wa	wi	we	wo
カ	キ	ク	ケ	コ	キャ	キュ	キェ	キョ	クァ	クィ	クェ	クォ
ka	ki	ku	ke	ko	kya	kyu	kye	kyo	kwa	kwi	kwe	kwo
サ	シ	ス	セ	ソ	シャ	シュ	シェ	ショ	スァ	スィ	スェ	スォ
sa	shi	su	se	so	sha	shu	she	sho	swa	swi	swe	swo
タ	チ	ツ	テ	ト	チャ	チュ	チェ	チョ	ツァ	ツィ	ツェ	ツォ
ta	chi	tsu	te	to	cha	chu	che	cho	tsa	tsi	tse	tso
ナ	ニ	ヌ	ネ	ノ	ニャ	ニュ	ニェ	ニョ	ヌァ	ヌィ	ヌェ	ヌォ
na	ni	nu	ne	no	nya	nyu	nye	nyo	nwa	nwi	nwe	nwo
ハ	ヒ	フ	ヘ	ホ	ヒャ	ヒュ	ヒェ	ヒョ	ファ	フィ	フェ	フォ
ha	hi	hu	he	ho	hya	hyu	hye	hyo	fa	fi	fe	fo
マ	ミ	ム	メ	モ	ミャ	ミュ	ミェ	ミョ	ムァ	ムィ	ムェ	ムォ
ma	mi	mu	me	mo	mya	myu	mye	myo	mwa	mwi	mwe	mwo
ラ	リ	ル	レ	ロ	リャ	リュ	リェ	リョ	ルァ	ルィ	ルェ	ルォ
ra	ri	ru	re	ro	rya	ryu	rye	ryo	rwa	rwi	rwe	rwo
ガ	ギ	グ	ゲ	ゴ	ギャ	ギュ	ギェ	ギョ	グァ	グィ	グェ	グォ
ga	gi	gu	ge	go	gya	gyu	gye	gyo	gwa	gwi	gwe	gwo
ザ	ジ	ズ	ゼ	ゾ	ジャ	ジュ	ジェ	ジョ	ズァ	ズィ	ズェ	ズォ
za	ji	zu	ze	zo	ja	ju	je	jo	zwa	zwi	zwe	zwo
ダ	ディ	ドゥ	デ	ド	デャ	デュ	ディェ	デョ	ドゥァ	ドゥィ	ドゥェ	ドゥォ
da	di	du	de	do	dya	dyu	dye	dyo	dwa	dwi	dwe	dwo
バ	ビ	ブ	ベ	ボ	ビャ	ビュ	ビェ	ビョ	ブァ	ブィ	ブェ	ブォ
ba	bi	bu	be	bo	bya	byu	bye	byo	bwa	bwi	bwe	bwo
パ	ピ	プ	ペ	ポ	ピャ	ピュ	ピェ	ピョ	プァ	プィ	プェ	プォ
pa	pi	pu	pe	po	pya	pyu	pye	pyo	pwa	pwi	pwe	pwo
	ティ	トゥ			テャ	テュ	ティェ	テョ	トゥァ	トゥィ	トゥェ	トゥォ
	ti	tu			tya	tyu	tye	tyo	twa	twi	twe	two
ヴァ	ヴィ	ヴ	ヴェ	ヴォ	ヴャ	ヴュ	ヴィェ	ヴョ	ヴゥァ	ヴゥィ	ヴゥェ	ヴゥォ
va	vi	vu	ve	vo	vya	vyu	vye	vyo	vwa	vwi	vwe	vwo
ガ゜	キ゜	ク゜	ケ゜	コ゜	キ゜ャ	キ゜ュ	キ゜ェ	キ゜ョ	グ゜ァ	グ゜ィ	グ゜ェ	グ゜ォ
nga	ngi	ngu	nge	ngo	ngya	ngyu	ngye	ngyo	ngwa	ngwi	ngwe	ngwo
	スィ				フャ	フュ	フィェ	フョ				
	si				fya	fyu	fye	fyo				
	ズィ											
	zi											
ン	ッ	ー	（	）								
nn	q	-	(	)								

鼻濁音：゜（JIS 0x212c）
促音：　ッ（JIS 0x2543）
長音：　ー（JIS 0x213c）
無声化・母音脱落：（（JIS 0x214a） および　） （JIS 0x214b）

［注1］　表記例のコードは全角片仮名（JIS X 0208-1983「情報交換用漢字符号系」）を使用するが，実際に使用するコード系まで規定するものではない。

［注2］　この表に定義されていない「ヂ」「ヅ」「ヲ」「ヰ」「ヱ」等の仮名記号は使用しない。

［注3］　長音記号として「ー（マイナス　JIS 0x215d）」は使用しない。

［注4］　網掛けされたアルファベット表記は発音の参考であり，音声認識詳細読み記号の規定対象ではない。

［注5］　半角片仮名で表記する場合，濁音では濁音記号「ﾞ」を使用する。

［注6］　半濁音では半濁音記号「ﾟ」を使用する。

こんにちは　こんにちわ (konnnichiwa)

学校へ　がっこうえ (gaqko-e, gaqkoue)

私を　わたしを (watashio)

私が　わたしが (watashiga, watashinga)

2.3 音声認識詳細読み記号に関する説明

主に開発者の使用を念頭においている。

＜規　定＞

● 「カタカナ」をベースに，表２に示す「音声認識詳細読み記号」を定める。

● 極力，発音どおりに記述すること。

● 鼻濁音の記号として「゜」を採用する。ただし，「カ゜」など表2にあげた形式での使用に限る。

● 長音化しないことを表す区切り記号は使用しない。

● 母音の無声化・脱落を表現するため括弧の記号を規定する。

＜補足説明＞

● 音声認識詳細読み記号は，音声合成読み記号との共通化を重視し，JEIDA 規格 日本語テキスト音声合成用記号の規格(JEIDA-62-2000) に掲載された表 2-1 を参考にした。

● 開発者が認識対象の発音をコントロールできることを重視した。ただし，各音声認識システムの性能向上のために，内部変換等の工夫をしてよい。

● 理解の手助けとして以下に例を示す。ここで（ ）に記した発音表記は，認識システムが読み記号を解釈し内部記号に置き換えた一例である。

（凡例）画面に表示する文字　読み記号の表記（ある認識システムで解釈された発音表記の例）

東京 トーキョー (to-kyo-)
丸の内 マルノウチ(marunouchi)
武蔵新城 ムサシシンジョー (musashishinnjo-)
大阪阿倍野橋　オーサカアベノバシ（o-sakaabenobashi）
経営　ケーエー (ke-e-)
経営　ケイエー (keie-)
私を　ワタシオ（watashio）
学校へ　ガッコーエ（gaqko-e）
こんにちは　コンニチワ（konnnichiwa）
私が　ワタシガ（watashiga）
私が　ワタシカ゜（watashinga）
デッドボール　デッドボール（deqdobo-ru）
デッドボール　デットボール（deqtobo-ru）
全員　ゼンイン（zenninn）
全員　ゼーイン（ze-inn）
秋田　ア（キ）タ (ak(i)ta)
私です　ワタシデ（ス）　(watashides(u))
ストップ　（ス）トッ（プ）　(s(u)toqp(u))
多い　オーイ（o-i）
氷　コーリ（ko-ri）
姉さん　ネーサン（ne-sann）
鼻血　ハナジ（hanaji）
続く　ツズク（tsuzuku）

2.4 全般的補足

● JEIDA の音声合成用読み記号は長音化を長音記号で表記するが，音声認識の場合は，エンドユーザが確実に長音化させて発声するか否かを予め記述することは難しい。これは，他の音韻についても同様であり，読み記号は音声合成用と音声認識用とでは，自ずからその役割・性質が異なる。

● 現行の各社 PC ソフトウエアの仕様は，一般的な「読み」であり，「とーきょー」「とおきょお」等は推奨されていない。

（例）ドラゴンスピーチ 6 片仮名で読み (トウキョウワ)
ViaVoiceV9 平仮名で読みと発音　(とうきょうわ)
SmartVoiceV4.0i 平仮名で読み　(とうきょうわ)
LaLaVoice2001 平仮名で読み　(とうきょうわ)

● 「音声認識読み記号」は，音声認識システムのサポートによって認識性能向上が期待できるが万能ではない。一方，「音声認識詳細読み記号」は，より正確に記述できる反面，間違って記述された場合，致命的になる可能性が高いことに留意する必要がある。

3. 音声認識・合成に係わる用語

音声認識や音声合成の開発者向けドキュメントは，特定の動作環境やエンジンの実装に依存して書かれていることが多く，用語とそれらの定義・関連に揺らぎがある。WG4 では，日本語の音声認識・音声合成エンジンに係わる用語を対象に，学会試行標準を検討している。標準化によって，システム開発者が既存のエンジンを使用する際に，動作環境やエンジンベンダーが変わっても違和感なく開発できることを目指す。

標準化のプロセスとして，まず日本国内で音声認識技術を製品化しているベンダー数社の現状調査を行った上で，実行可能性のある標準化案を小委員会内で検討している。主な方針は以下のとおりである。

● 既存の標準を尊重する（JEITA,W3C 等）

● 関連分野のデファクトを尊重する（かな漢字変換等）

● 専門書や研究論文等の定義や慣用的用法を尊重する

● 現状調査（既存の製品仕様やドキュメント）を考慮する

今後メーリングリスト等を通じて，素案に関するアンケートを行う予定である。

3.1 標準化の対象と目標

開発者やユーザにとって用語の不統一が混乱を招きやすいと思われる事項で，かつ，統一化の合意が得られやすい提案に限定する。また，既存の記述の変更を求めるものではなく，

開発者間のコミュニケーションにおいて曖昧性が生じやすい場合に推奨する用語を提案する。

用語の選定に当たっては，用語間の階層関係（依存関係）を考慮するなどにより用語の位置づけを明確にすること，および用語がどのような状況で用いられどのように役立つべきかを明確にすることを目指す。

エンドユーザが製品を通じて目にする用語には，様々な不統一が存在している。しかし既に製品として出荷されていることもあり，エンドユーザが慣れ親しんだ用語を変更することには問題も多い。そこで現時点では，エンドユーザ向け用語の統一案は作成せずに各社の判断にゆだねる。

一方，開発者向け用語については，以下のような目標を設定して用語標準化を提案する。

● 音声認識・音声合成エンジンの機能や性能を明確に記述でき，開発者がエンジンを選択する際の補助となること

● システムに組み込まれた特定のエンジンを，同等の機能・性能を持つ他のエンジンと容易に交換できること

これらの目標が達成されることで，エンジンの開発や改良，および応用システムの開発が促進されることを期待している。

3.2 音声に関する用語

「音声」の用法には，広く音響信号を示す場合と，言語情報を伴う音に意味を限定する場合がある。特に，音声認識や音声合成が応用システムの一部として用いられる場合には，これらの概念を明確にしておくことが重要である。

入力音声

音声認識システムの入力となる音を総称して「音響信号」と呼ぶ。「音響信号」の内容は「音声信号」と「雑音信号」に分離される。また，音響信号の時系列情報から「休止区間」を取り除いて「発話区間」を得ることを（発話区間の）「切り出し」と呼ぶ。

出力音声

音声合成システムの出力となる音を総称して「音響信号」と呼ぶ。「音響信号」には，「録音音声」および「合成音声」のほかに楽音や効果音などが含まれる。テキスト情報から「合成音声」を得る手法を「テキスト音声合成」と呼ぶ。「テキスト音声合成」において埋め込まれる休止を「ポーズ」と呼ぶ。

なお「プロンプト」は，アプリケーション中で音声を利用する方法のひとつを示した用語であり，上に定義した用語群とは区別して使用すべきである。

3.3 音声認識に関する用語

音声認識手法

「連続音声認識」では，用途を表す用語と音声認識エンジンの手法そのものとが混同されがちである。そこで連続音声認識の手法として次の二つの分類を使用する。

A: 統計的言語モデル(n-gram)を用いた連続音声認識

B: 構文制約情報を用いた連続音声認識

いずれかの手法を限定したい場合は明示的に記述する。

例: Julius は統計的言語モデルを用いた「連続音声認識」を行っている。

例：Julian は構文制約情報を用いた「連続音声認識」を行っている。

なお，「ディクテーション」「ナビゲーション」「音声コマンド」などは音声認識手法ではなくアプリケーションを示す用語であり，手法を示す際には用いない。

音声認識性能

音声認識エンジンの性能の求め方や結果の表示方法は統一されていない。一般に性能とは認識率を指すが，処理速度が重視される場合もある。

アプリケーション開発者が目的に合った性能のエンジンを選ぶには，性能を記述する用語とその定義を標準化する必要がある。

アラインメント作業が必要な場合の音声認識性能は，「正解精度」(Accuracy) もしくは「誤り率」(Error Rate)を用いて示すことを推奨する。

正解精度 = (正解数 - 挿入数)×100 / 入力総数

誤り率 = 100 – 正解精度

1 発話で 1 タスクが達成される場合も，挿入・脱落はないものとして，「正解精度」もしくは「誤り率」を用いて示す。

「正解精度」や「誤り率」の値を示す場合は単位を明記して「文字正解精度」「単語正解精度」「音韻正解精度」のように用いる。

音声認識応用システム全体の性能は，上記定義の他にも発話者の声質や利用環境，ユーザインタフェース設計などに大きく依存する。将来的には，エンジン性能の評価尺度をさまざまな観点から明らかにし，必要な情報の整備を促したい。

3.4 音声合成に関する用語

3.4.1 音声合成制御

テキスト音声合成の制御機能を表す用語は，様々なものが用いられている。また音声合成アプリケーションの開発者を対象とした標準やデファクトとして，JEIDA 日本語テキスト

音声合成用記号，W3C Aural CSS，W3C SSML，Microsoft speech API などが策定されている。

一方，個々のエンジンは既存の標準に従いつつ，独自の解釈や拡張が施されている場合も多い。機能ごとに設定される値は，物理量などの客観的な値ではなく相対値を用いる場合が多く，エンジンが API によって抽象化されていても実質的には互換性が妨げられている。

以下では標準案の候補となる用語とその定義を示す。

合成音声の速さ

「速さ」「速度」「話速」は合成音声の速さを示す値であり，速いほど大きい値として定義される。「速さ」の例として，1 秒あたりのモーラ数や 1 分当たりの単語数などがある。

英語等では「話速」として 1 分当たりの単語数[Words per minute, WPM]を用いることが多く，Aural CSS (W3C [4])の speech-rate 属性もそのように定義されている。マルチリンガルの音声システムを実現する上では，言語によらず統一的に話速を表現できることが望ましい。

声の高さ

「高さ」「ピッチ」「基本周波数」は，合成音声の声の高さをヘルツ[Hz]で示した値である。Aural CSS (W3C)の pitch 属性も周波数[Hz]として定義されている。ただし値の用法は実装依存となる。実際のアプリケーションでは，規定値からの相対指定（例：75%, 150%など）の方が便利なため多く用いられている。

声の大きさ

「音量」「ボリューム」「大きさ」は声の大きさを表す値である。物理量としては音圧レベル[dBSPL]に対応する。しかし既存の実装においては，0 を最小値とし，特定の値を最大値とした相対値で振幅を決めている場合が多い。

Aural CSS (W3C)の volume 属性は 0-100 の値を取ると定義しているが，こうした値は環境に応じて制御されるものであり，同じ volume 値であっても騒音レベルによって異なる振幅となるなど，アプリケーションの観点から与えた定義をそのままエンジン制御に適用するのは適さない。

声　質

異なる声質の合成音声を使い分け，GUI における色やフォントのように利用することが考えられる。「音声フォント」「キャラクタ」「ボイスフォント」「声質」「バリエーション」は声質の識別子の呼称である。一般に「音声フォント」には人名など任意の文字列を使用できる。

Aural CSS (W3C)は voice-family 属性を提供しており，male, female, child など一般的な呼称と，特定のインスタンスを表す呼称を併用できる。

3.4.2 音声合成の手法および性能

アプリケーション開発者は，音声合成エンジンを選ぶ際に，エンジンが想定している音質レベルを知る必要がある。

音声符号化の分野では CD 相当(約 44KHz, 16bit)，FM ラジオ相当(約 22KHz, 16bit)，電話相当(約 8KHz, 8bit)など既存メディアを目安にする方法があるが，合成音声の品質は帯域やビット数以外に，手法や実装の違いによっても異なる。また，特に視覚障害者向けアプリケーションでは速い話速が要求されることが多く，実用的な話速の最高値などもエンジン選びの指標となり得る。

一方，テキスト音声合成では，読み付与の精度も重要な評価尺度となる。ここでは 2 つの性能評価尺度を示す。

「読み付与精度」：仮名漢字混じり文から読みを生成した際の単語単位の精度（計算方法は音声認識の単語認識精度に準ずる）

「正聴率」：被験者に合成音声を聴取させ書き起こしを行った際に，正しく単語を聞き取れた割合

正聴率＝正解音節数／全音節数

評価方法の詳細は JEITA での標準化活動を参考にする。

3. おわりに

試行標準は，今後，メーリングリスト等を通して関連分野の研究者・技術者の方々の意見を反映させた後，情報処理学会の Web 上で順次公開し，「役立つ試行標準」の提供を行っていきたい。

参考文献

[1] 新田ほか：音声言語情報処理に関する情報処理学会の試行標準策定活動，情報処理学会研究報告，SLP-40-10, pp.57-60 (2002-02).

[2] 小泉保「日本語の正書法」，大修館書店

[3] 国語審議会「外来語の表記」
http://www.konan-wu.ac.jp/~kikuchi/kanji/gairai.html

[4] Aural style sheets,
http://www.w3.org/TR/REC-CSS2/aural.html