日本音響学会誌 57 巻 12 号 (2001), pp. 776-782

解説

音声合成用記号の標準化について*

赤羽 誠 (ソニー㈱)**・蓑輪利光 (松下通信工業㈱)***・板橋秀一 (筑波大学)****

43.15.+s; 43.72.Kb

1. はじめに

近年，合成音声の音質向上，及び，日本語処理技術の進展に伴うテキストの読み精度の向上により音声合成，特に，テキスト音声合成を使用したアプリケーションやサービスが広まってきている。このような中で各メーカが提供する音声合成装置，サービスあるいはオペレーティングシステム (OS) の応用プログラムインタフェース (API) レベルにおいて，音声合成システムのインタフェース，特に，音声合成用記号が異なるという状況が出現してきた。音声合成用記号は，標準化されたものがあれば，ユーザやサービス提供者が独自に決める必要のないものであり，また，音声合成装置を提供するメーカ側から見てもアプリケーション・サービスによって異なる複数の記号をサポートする必要がなくなる。

一方，ヨーロッパではESPRIT Speech Assessment Methods (SAM) プロジェクト[1]においてヨーロッパ系言語のコンピュータで処理可能な音声記号の検討が行われ，その後，多くの言語に拡張が行われているが，日本語は検討対象になっていない。更に，音声入出力用のAPIもいろいろなプラットホームで検討されているが，日本語特有の表記が考慮されていないなどの状況が見られるようになった。

現在，日本において標準と呼ばれるものには，2000年3月に㈱日本電子工業振興協会 (Japan Electronic Industry Development Association: JEIDA) (脚注1) において制定されたJEIDA-62-2000「日本語テキスト音声合成用記号の規格」[2]がある。また，マイクロソフト社のSpeech API[3](以後，SAPIと省略する)，Voice XML[4]などで，音声合成用記号が取り扱われるようになってきた。

本解説ではJEIDA規格「日本語テキスト音声合成用記号の規格」を紹介すると共に他の標準化動向との関係を説明する。

2. 日本語テキスト音声合成用記号の規格 [2]

2.1 目　　的

日本語テキスト音声合成を利用した各種のアプリケーション及びサービスで共通に利用可能な日本語テキスト音声合成用記号を定め，システム開発者とユーザの便宜を図ることによりテキスト音声合成技術のよりいっそうの利用拡大を図ることを目的としている。

2.2 概念と分類

日本語入力を対象として，日本語の特性を考慮して記号を決定した。ただし，音声合成用記号によって，出力される合成音声そのものを規定するのは困難であり，出力される音声は各音声合成システムにより異なることになる。更に標準化においては下記の点に留意している。

1) 特定のアプリケーションやプラットホーム (ハードウェア・アーキテクチャ，オペレーティングシステム，プログラミング言語，文字コード系) などに依存しないこと

* The standardization of symbols for speech synthesizer.

** Makoto Akabane (Sony Corporation, Tokyo, 108-6201)

*** Toshimitsu Minowa (Matsushita Communication Industrial Co., Ltd., Yokohama, 224-8539)

**** Shuichi Itahashi (University of Tsukuba, Tsukuba, 305-8573)

(脚注1)㈱日本電子工業振興協会 (JEIDA) は，2000年11月，㈱日本電子機械工業会 (Electronic Industries Association of Japan: EIAJ) と統合され，㈱電子情報技術産業協会 (Japan Electronics and Information Technology Industries Association: JEITA) となり，音声入出力方式標準化委員会において引き続き音声技術の標準化の検討を行っている。

2） 広範なアプリケーションや各種プラットホームに共通して使える汎用性を持たせること

一方，テキスト音声合成装置のインタフェースにおいては，表記方法以外に音声合成装置を制御する部分も含まれるが，これらはアプリケーションあるいはプラットホームに固有なものが多く，各アプリケーションあるいはサービスにおいて個別に対処すればよい問題として本規格には含まれていない。従って，本規格はテキスト音声合成装置のインタフェース全体の仕様ではなく，音声合成用記号の表記方法のみを対象としている。

本規格では，音声合成用記号の持つ情報（読み，韻律，制御情報）と階層（テキスト，仮名，異音）の二つの要因を整理し，**表-1** のように音声合成用記号の分類・定義を行い，テキスト音声合成用記号を，音声合成用記号の持つ情報からの「読み記号」「韻律記号」「テキスト埋め込み制御タグ」と，階層からの「仮名レベルの表記」「異音レベルの表記」「テキスト埋め込み制御タグ」に分類している。この分類とテキスト音声合成処理の関係を**図-1** に示す。

表-1　テキスト音声合成用記号の分類

階層	読み情報	韻律情報	制御情報
テキストレベル	—	—	テキスト埋め込み制御タグ
仮名レベル	仮名レベルの表記		
	読み記号	韻律記号	
異音レベル	異音レベルの表記		
	読み記号	韻律記号	

（1） テキスト音声合成システムへの入力となるテキストレベル

テキストレベルの表記はテキスト音声合成システムへの入力となるもので，その中で音声合成部を制御するためのテキスト埋め込み制御タグを規定した。テキスト埋め込み制御タグは世の中の標準化動向に合わせてXML形式を採用した。また，テキストレベルだけではなく，仮名レベル，異音レベルでの表記にも使用可能なものとしている。

（2） 日本語解析処理からの出力となる仮名レベル

仮名レベルは日本語テキスト解析部の出力であり，規則音声合成部の入力となるもので，概ね音素表記に相当するものとして考えられる。このレベルは主に仮名で表記できるため「仮名レベル」と名づけた。仮名レベルの記号は読み記号と韻律記号及びテキスト埋め込み制御タグから構成される。

（3） 仮名レベルで表記できない異音までを含めた単音（phone）を記述できる異音レベル

異音レベルは仮名レベルよりも詳細な音を表現できる規則音声合成部の入力で，概ね音声表記に相当するものとして考えられ，以下の観点からまとめた。

1） 仮名レベルの読み記号に相当する日本語の単音節を1種類の代表的な国際音声記号（International Phonetic Alphabet：IPA）で表記する。更に，母音の無声化など音環境

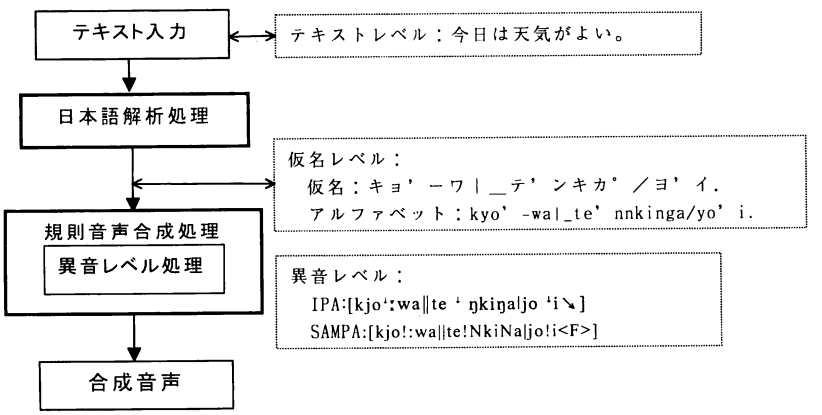

図-1　テキスト音声合成処理と記号の関係

により変わり，単音節で表記できない異音を表記する。

2） 仮名レベルの韻律記号に相当する韻律の表記をIPAをベースに行う。

また，IPAで日本語を表記することにより多言語の中での日本語の表記方法を明らかにし，国際的な記号の標準化の動向に対応する。異音レベルの記号はIPA及びヨーロッパで標準化の進んでいるSAM（Speech Assessment Methods）コンソーシアムのSAMPA（Speech Assessment Methods Phonetic Alphabet）[1,5]，特に拡張SAMPA（X-SAMPA）[6]に準じたアルファベット（ASCIIコード）で表記する。

2.3 仮名レベルの表記

2.3.1 読み記号

読み記号は平成2年度にJEIDAが提案した音声表記用記号表[7]をベースにしている。1単音節1表記を原則に，撥音，促音，長音を含む。読み記号表を**表-2**に示す。使用にあたっての留意点は下記のとおりである。

（1） 文字コード系は規定しない

片仮名とアルファベットの表記を行い文字種は規定するが，アプリケーションプラットホーム等に依存する文字コード系は規定していない。

（2） 鼻濁音の表記はあるが無声音の表記はない

鼻濁音はテキスト解析処理でなければ鼻濁音と濁音の判断ができないので，表記方法を規定したが，母音の無声化については前後の音素環境が分かれば規則音声合成処理過程で判断可能なため規定していない。

（3） 正書法とは必ずしも一致しない

1単音節1表記を原則とするため，「ヂ」「ヅ」「ヲ」「ヰ」「ヱ」等は同じ合成音となる「ジ」「ズ」「オ」「イ」「エ」等に統一した。

2.3.2 韻律記号

アクセントは東京アクセントの表現を行うことを前提にした表記方法を採用した。また，副次アクセントなど微妙な表現を必要とされる場合も想定されたが，複雑な表記方法を採用するとアプリケーションでの利用が難しくなることから，できるだけ簡略化したものとした。更に詳しい表記のためには異音レベルの表記を使用することになる。**表-3**に韻律記号を示す。

2.3.3 制御情報

初期の検討案[7-10]においては，コメントの開始・終了，制御情報の開始・終了の記号を用意する方向で検討を行っていたが，制御タグのXML化に伴い，XMLの表現に合わせ，制御情報を仮名レベルの表記として特別に表すことは行わなかった。

2.4 異音レベルの表記

2.4.1 読み記号

仮名レベルよりも細かな読みを指定するために，日本語の単音節をIPAで表記した。しかし，今のところ，音声学の分野でもIPAを使用した日本語の標準的な表記方法は存在せず，いろいろなIPAが使用されて日本語が表現されている。そこで，工学的な見地から，まず，仮名レベルの読み記号を一つのIPAで表すことを行った。参考にしたのが日本音声学会でのIPAワーキンググループの簡略音声表記の研究調査[11]である。この調査結果で，単音節を単独で発声したとき表記例の一番多かったIPA表記を採用し，後から全体としての整合性をとっている。

2.4.2 IPAのASCIIコード表記

IPAの記号は特殊なフォントを使用し，今のところ多くのアプリケーションで利用するのが困難な状況である。そこで，IPAをコンピュータで取り扱い易い記号で表す必要がある。本規格では，初期の検討段階においてWorldbet[12]を使用していたが，最終的にはヨーロッパで標準化の進んでいるSAMPA[1,5]を採用した。

2.4.3 異音レベルの読み記号表

異音レベルと仮名レベルの表記は1対1に対応しているわけではないが，**表-4**に異音レベルで使用している記号を仮名レベルのアルファベット表記に対応させて示す。異音レベルの読み記号表は仮名レベルの読み記号表にこれらの記号が追加され，構成されている。

2.4.4 異音の表記

単音節の単独の発声では表せない音を「異音の表記」として定義する。どこまでを異音として表記するかは，議論の分かれるところであるが，日本音声学会IPAワーキンググループの研究調査を参考にして，10種類の表記すべき異音を選択した。**表-5**に表記を示す。モーラ境界は異音ではないが，異音レベルで必要な表記として採用し

表-2 仮名レベルの読み記号

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

鼻濁音：゜ (JIS 0×212c)
促　音：ッ (JIS 0×2543)
長　音：ー (JIS 0×213c)

ている。

2.4.5 韻律記号

異音レベルの韻律記号でも読み記号と同様に仮名レベルの表記をIPAで表すことを行った。ただし，IPAは韻律の表現まで含めて表記をすることを主な目的としていないため，不十分なところも見受けられるが，IPAのSUPRASEGMENTALSの記号を使用して表記した。採用された記号を表-4に示す。今後，更にアプリケーションの動向も見ながら，使い易い記号の検討が必要とされる。

表-3 仮名レベルの韻律記号

	韻律記号	読み方	JISコード	ASCIIコード
アクセントの位置	'	アポストロフィー，シングルクオート	0×2147	039
アクセント句の区切り	／	スラッシュ	0×213f	047
フレーズの区切り	|	タテボウ	0×2143	124
文末（通常のイントネーション）	.	ピリオド	0×2125	046
文末（疑問のイントネーション）	?	クェッションマーク	0×2129	063
文末（驚きのイントネーション）	!	エクスクラメーションマーク	0×212a	033
1モーラのポーズ	_	アンダースコア	0×2132	095

表-4 異音レベルで使用している音声記号

仮名レベル	IPA	IPA Number	SAMPA	備考	仮名レベル	IPA	IPA Number	SAMPA	備考
a	a	304	a		h	ç	138	C	ヒで使用
i	i	301	i		v	v	129	v	
u	ɯ	316	M		—	z	133	z	
e	e	302	e		h	h	146	h	
o	o	307	o		—	ɦ	147	h\	有声声門摩擦音
—	˞	402 A	_0	無声化	ts	ts	211	ts	
—	:	503	:	長音化	ch	tʃ	213	tS	
y	j	153	j		z	dz	212	dz	
w	w	170	w		j	dʒ	214	dZ	
w	ʷ	420	_w		m	m	114	m	
p	p	101	p		n	n	116	n	
t	t	103	t		ny	ɲ	118	J	
k	k	109	k		ng	ŋ	119	N	
b	b	102	b		—	ɴ	120	N\	
d	d	104	d		—	˜	424	~	鼻音化
g	g	110	g		—	.	506	.	Syllable break
—	β	127	B	有声両唇摩擦音	—	ˈ	518	^	アクセント（上がり）
r	ɾ	124	4		—	ˌ	517	!	アクセント（下がり）
f	Φ	126	p\		—	|	507	|	アクセント句の区切り
f	f	128	f		—	‖	508	‖	フレーズの区切り
s	s	132	s		—	↘	511	<F>	文末（通常のイントネーション）
sh	ʃ	134	S		—	↗	510	<R>	文末（疑問のイントネーション）

2.5 テキスト埋め込み制御タグ

テキスト埋め込み制御タグ（以後，制御タグと省略する）の記述フォーマットはXMLに準拠している。その定義においては，各社の音声合成装置の仕様，マイクロソフト社のSAPI[3]，W3C（World Wide Web Consortium）Aural Style Sheet[13]などを参考にしながら，検討を行った。

最終的に，SAPIで定義されている内容を考慮し，日本語及び日本語音声合成器の特徴も考慮に入れて，必要最低限の制御タグを選定した。

2.5.1 テキスト埋め込み制御タグの分類

制御タグは下記のように，直接，音声合成システムを制御するもの，読み方を指定するもの，主にテキスト解析精度を向上させるものとに分類することができる。読み方を指定するタグでは，出力される音声は変化するが，テキスト解析精度向上のためのタグでは，出力される音声の変化はそれぞれの処理系に依存することになる。

（1） 音声合成システム全体の制御

ブックマーク（BOOKMARK），スピーチ（SPEECH），言語指定（LANG），音声フォント指定（VOICE），リセット（RESET）

表-5 異音の表記

異音	IPA 表記	SAMPA 表記	表記例（「日本語」[IPA] [SAMPA]）
母音間の有声声門摩擦音	ɦ	h\	「気配」[ke ɦ ɑi] [keh\ai]
母音の無声化	◌̥	_0	「岸」[kʲi̥] [ki_0Si]
母音間の破擦音の摩擦音化	z	z	「アザ」[aza] [aza]
撥音の後続音への同化	ɴmnɲŋ͡	N\ mnJN～	「散歩」[sampo] [sampo]
促音の後続音への同化	同じ子音の記号を二つ続ける	同じ子音の記号を二つ続ける	「サッカー」[sakka:] [sakka:]
長母音	:	:（コロン）	「アー」[a:] [a:]
二重母音	特に用意しないが，表記例の区別を行う	特に用意しないが，表記例の区別を行う	「問う」[tou̯] [toM]/[[too] [too]/[to:] [to:]
母音間の有声破裂音の摩擦音化	β	B	「アブブ」[aβu̥βu̥] [aBMBM]
半母音	拗音 j，合拗音 w，円唇化 ʷ	拗音 j，合拗音 w，円唇化_w	「休暇」[kju̥:ka] [kjM:ka]
モーラの境界	.	.（ピリオド）	

（2） 読み方の指定

ポーズ（SILENCE），強調（EMPH），綴り読み指定（SPELL），発音読み指定（PRON），話速(RATE)，音量(VOLUME)，ピッチ(PITCH)

（3） テキスト解析精度向上

品詞指定（PARTOFSP），内容指定（CONTEXT），単語登録（REGWORD）

2.5.2 相対値指定と絶対値指定

各社の日本語音声合成システムにおいて，話速，音量，ピッチをレベルによって指定する方法が多く見られたので，タグの属性に LEVEL を用意してレベル指定可能にした。合わせて，レベル指定に絶対的に指定する方法と相対的に指定する方法を用意した。

3. 他の音声表記との関係と標準化動向

JEIDA 規格の仮名レベルの読み記号と韻律記号は音声合成装置，アプリケーション，マイクロソフト社の SAPI での採用例も見られる。更に，異音レベルの表記で SAMPA を使用したため，日本語の表記方法として広まっていくことが期待される。

一方，テキスト埋め込み制御タグでは，XML をベースにした表記方法が主流となることが予想され，その結果，VoiceXML，W3C[14]の動向が規格に大きく影響してくると思われる。現在は JSML（Java Speech ML）を推奨している VoiceXML も，W3C が策定している JSML をベースの SSML（Speech Synthesis ML）[15]の方向に行くものとの思われる。

4. おわりに

JEIDA 規格においては，テキスト音声合成で，規格化し易く，かつ，必要とされている音声合成用記号の規格化を行った。ただし，本規格は最低限必要なものを規定したに過ぎず，今後，音声合成技術やその応用の発展や他の標準化動向に合わせ，改良，更新が継続される必要がある。

謝辞

「テキスト音声合成用記号の規格」は，㈳日本電子工業振興協会 音声入出力方式専門委員会において平成 12 年度に JEIDA 規格として制定されるまでご検討をいただいた歴代委員，音声合成 WG メンバー，事務局の関係各位，平成 13 年度からの㈳電子情報技術産業協会 音声入出力標準化委員会，音声合成 WG メンバー，事務局の関係各位の成果であることを記します。

文献

[1] D. Gibbon, R. Moore and R. Winski, Ed., *Handbook of Standards and Resources for Spoken Language Systems* (Mouton de Gruyter, Berlin/New York, 1997).

[2] ㈳日本電子工業振興協会, "JEIDA 規格「日本語テキスト音声合成用記号の規格」," JEIDA-62-2000 (2000).

[3] Recent Speech Developments at Microsoft,

http: //microsoft.com/speech/,
http: //research.microsoft.com/stg/.
[4] Voice XML, http: //www. voicexml. org/.
[5] SAMPA, http: //www.phon.ucl.ac.uk/home/sampa/home.htm.
[6] X-SAMPA, http: //www.phon.ucl.ac.uk/home/sampa/x-sampa.htm.
[7] ㈳日本電子工業振興協会編，ヒューマンメディア情報処理の標準化に関する調査研究報告書（㈳日本電子工業振興協会，東京，1996），pp. 179-188.
[8] ㈳日本電子工業振興協会編，ヒューマンメディア情報処理の標準化に関する調査研究報告書（㈳日本電子工業振興協会，東京，1997），pp. 289-290.
[9] ㈳日本電子工業振興協会編，ヒューマンメディア情報処理の標準化に関する調査研究報告書（㈳日本電子工業振興協会，東京，1998），pp. 113-120, 170-175.
[10] ㈳日本電子工業振興協会編，ヒューマンメディア情報処理の標準化に関する調査研究報告書（㈳日本電子工業振興協会，東京，1999），pp. 115-119, 161-174.
[11] 大西雅行，土岐 哲，堺辻正剛，“日本語の音声表記法調査，”1995年（平成7年）度日本音声学会全国大会予稿集，11-17（1995）.
[12] J.L. Hieronymus, *ASCII Phonetic Symbols for the World's Languages: Worldbet* (Technical Report, AT & T Bell Labs., Murray Hill, 1994).
[13] W3C, Cascading Style Sheets, level 2, http: //www.w3.org/TR/WD-CSS2/.
[14] W3C Voice Browser, http: //www.w3.org/Voice/.
[15] Speech Synthesis Markup Language Specification, http: //www.w3.org/TR/2001/WD-speech-synthesis-20010103/.

赤羽 誠

1977年，早稲田大学理工学部電気工学科卒。1979年，早稲田大学大学院修士課程修了。同年ソニー㈱入社。技術研究所，情報通信研究所，D21ラボラトリー等で音声合成，音声認識の研究開発及び商品化に従事。現在，同社AV/IT開発本部勤務。日本音響学会，電子情報通信学会，人工知能学会各会員。

蓑輪 利光

1981年，早稲田大学理工学部電気工学科卒。1983年，早稲田大学大学院修士課程修了。同年松下電器産業㈱入社。松下通信工業㈱技術本部開発研究所，同AV & C研究所等で音声合成，音声認識の研究開発及び商品化に従事。現在，同社技術本部マルチメディアソリューション研究所勤務。日本音響学会，電子情報通信学会会員。

板橋 秀一

昭和39年東北大・工・通信卒。昭45同大学院(博)修了。同年東北大通研助手。昭47電子技術総合研究所入所。昭52-53ストックホルム王立工科大学客員研究員。昭57筑波大電子・情報工学系助教授。現在同教授。工博。音声・画像・自然言語処理の研究に従事。平11～12年音声研究委員会委員長。電子情報通信学会，情報処理学会，人工知能学会，言語処理学会，認知科学会，IEEE，ASA，ESCA各会員。