解　説

自然言語処理技術の応用

6. 音声合成における自然言語処理†

匂　坂　芳　典††

1. はじめに

音声は，人間どうしの間，人間・機械間の情報伝達に手軽で便利な手段であり，音声合成の分野では，任意の伝達内容を自然な音声で出力することを目指した研究がすすめられてきた．伝達内容の音声を出力するための入力情報としては，通常の文章テキストがまず自然に考えられる．実際，文章テキストを入力として音声出力を行う「テキストからの音声合成」の研究は精力的にすすめられ，合成音声の品質も大きく向上し，一部では実用に供する段階にまできている．しかし，文章テキストは人間にとって便利な一つの入力インタフェースではあるが，音声合成にとっては必ずしも都合のよい情報形態ではない．音声合成に必要な情報をテキストから抽出するためには究極的には文章内容の理解が必要であり，文章テキストを入力とすることは，自然言語処理がかかえる本質的な問題をそのまま引き継ぐことになる．さらに加えて，音声合成の技術が広まるにつれて，その適用領域も単なる音声出力による情報伝達といった当初の目的から発展し，より状況にあった多様な合成音声の生成が求められている．このような適用領域の拡大にともなって，テキストだけでは規定できない音声発話のための付加入力情報の必要性も指摘されている．

ここではまず次章で，日本語テキストを入力とした音声合成について解説し，音声合成で必要な情報をテキストから得るための自然言語処理を紹介する．**3.**, **4.** では，合成音声に自然なイントネーション，リズムを与える上で言語情報がどのように反映されているかをさらに詳しく紹介する．最後に **5.** では，音声を合成するために必要な入力言語情報のあり方について述べる．

2. テキストからの音声合成

図-1 に日本語テキストからの音声合成システムの概要を示す．まず，入力された仮名漢字混じり文章で表された日本語文章テキストをもとに，構成単語の同定，局所的な単語・文節間の依存関係の分析といった文章解析（主に形態素解析）がなされる．次に，これによって得られた単語の読み，アクセント，品詞といった語彙情報や，単語・文節間の局所的な依存構造を入力として，音韻規則，韻律規則が適用され，音声の発音表記や各音韻の継続時間長，文節ごとのアクセントの位置，大きさ(ストレス・レベル)，発話境界位置，境界におけるポーズの有無・長さなどのいわゆる韻律制御パラメータが決定される．最後に，音声

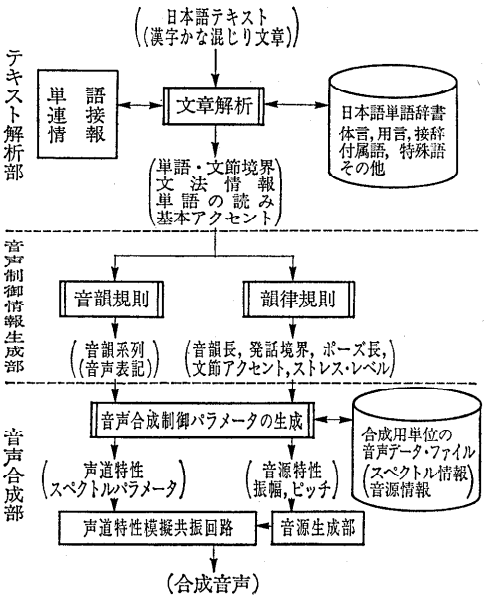


図-1　日本語テキストからの音声合成システム

† Natural Language Processing in Speech Synthesis by Yoshinori SAGISAKA (ATR Interpreting Telecommunications Research Laboratories).

†† ATR 音声翻訳通信研究所

合成部では得られた発音表記をもとに最適な合成単位のスペクトル・パラメータが音声単位ファイルから選び出され，音韻長データに基づいた合成単位の伸縮結合がなされ，これに基づいて声道特性の制御がなされる．また，韻律制御パラメータをもとに，基本周波数の時間変化パタンが生成され，これに基づいて駆動信号が生成され，声道特性を模擬する共振回路により音色が重畳され，合成音声が得られる．

以下，上記の音声合成処理の中で特にテキストを入力としたことにともなって生ずる，文章解析（形態素解析），読み付け処理（音韻規則），アクセント処理（韻律規則）について述べる．

2.1 文章解析

テキスト中の単語の読みやアクセントを知るためには形態素解析が必要であるが，よく知られるように日本語では単語単位の分かち書きの習慣がない上，いわゆる漢字かな表記の自由度が高いため，単語の同定が難しい．この形態素解析の問題は，本特集の他の応用にも関連し，そちらで詳しく説明されると思われるが，これまでに最長一致選択，文節数最小化といった種々のヒューリスティクスをもとに複数の単語候補のもっともらしさを評価して選択する方法[1),2)]や単語間の意味関係を用いた処理[3),4)]，単語連接の統計的な性質を用いた分割法[5)]などいろいろ提案されてきた．これらの方法はそれなりに有効であり，実際のシステムでもこのような方法が使用されていると思われる．しかし，形態素解析の問題は，究極的には意味理解を含む高度な知識や解析を必要とするため，完全な解決は難しい．このため，文章解析はいわば，テキストからの音声合成の泣き所であり，現在のテキスト音声合成システムでは，ポストエディティングを可能にするなど文章解析によって生ずる単語同定誤りに対処するための工夫がなされている．

この形態素解析の難しさは，文章テキストが音声合成システムへの最適な入力情報形態でないことを示している．研究としては，人間とのインタフェースを重視し，今のままの文章テキストを対象として理想的な解析手法，文法，語彙辞書をさらに追求する方向が考えられる．この研究が進展すれば，現在の合成システムにきちんとした文章構造の解析を取り入れることが可能で，後の章で述べるような合成に必要な情報を得ることが期待できる．しかし，また一方，とりあえず音声出力するために若干の不便を人間の側に強いる折衷案的な解決の方向も十分考えられる．すなわち，音声合成機能の実際的な利用を考える上では，ちょうど日本語ワードプロセッサへの入力方法がそうであったように，必要に応じて合成に必要な情報を人間がテキストとともに入力することで，技術的な難しさを回避できる．このような方法をとる場合は，人間があまり無理なく適応，学習できる入力形式，操作手順の模索が必須であると思われる．

2.2 読み付け処理

日本語の場合は音韻の連接によって生ずるスペクトルの変化が英語などに比べて小さいため，辞書にある単語の読みをもとにして合成に必要な音韻に対応する音声単位スペクトルを得ることができる．しかし，次のような場合は辞書に記載されている読みの変形，複数の読みからの選択などの処理が必要である．

（1） 連　濁

青山＋通り（とおり）→青山通り（あおやまどおり）

「本」一本（ぽん），二本（ほん），三本（ぼん）

（2） 接辞，造語成分の読み分け

「無」無理解（む），無遠慮（ぶ）

「人」日本人（じん），見物人（にん），さすらい人（びと）

（3） 固有名詞などの読み分け

清水（しみず，きよみず，せいすい）

靖（やすし，きよし，おさむ，せい）

（4） 略語，意味表記としての漢字使用

参院予算委（さんいんよさんい）

大人，中人，小人（だいにん（おとな），ちゅうにん，しょうにん（しょうじん））

上記の問題中，（1），（2）のある程度は規則で対処可能であり[6),7)]，他の問題も，漢字にルビをふるのと同様に適切な読みやアクセントの補助記号を取りいれることでかなり解決されると思われる．これらの本質的な解決には，前節で述べた，豊富な知識，辞書と精緻な文章解析技術が必要であることはいうまでもないが，（4）の例で代表される読み付けの問題は別の問題の所在も示している．これらの例は，意味伝達を目的とした書き言葉としての漢字表記であって，読むことを意識し

表-1 単語，文節の結合にともなうアクセント変化規則

連接種類		アクセント結合様式		単語例	結合例
文節内結合	付属語結合	従属型		た，ほど	さがした およいだ
		不完全支配型		らしい， ので	さがすので， およぐので
		支配型		まい，ます	さがしません， およぎません
		融合型		せる，れる	さがせる， およがせる
	複合語結合	接辞結合	標準型	質，官，学	キンゾクシツ (金属質)
			平板化型	的，化，性	キンゾクテキ (金属的)
		自立語結合	保存型	頭高・中高型名詞	キンゾク カゴウブツ (金属化合物)
			生起型	平板・尾高型名詞	キンゾク コウタク (金属光沢)
文節間結合		先行単語優先型		シロイ＋クモ →シロイクモ(白い雲)	

た文章テキストではない．音声合成システムの目的が音声による自然な情報伝達であるとすれば，辞書の読みどおりに発音することがかならずしも良いとは限らない．これらは，書き言葉の内容を適切な話し言葉に変換するといった，新たな自然言語処理の取組みの必要性を示している．

2.3 アクセント処理

日本語では，単語の連接によって次のようなアクセントの移動，生成，消失が起きる．

おこ┐る＋られる→おこられ┐る(怒られる)　(移動)

ごうせい＋き→ごうせ┐いき(合成器)　(生成)

たべ┐る＋だけ→たべるだけ(食べるだけ)(消失)

このアクセント変化は規則的であり，多くの場合，複合語・文節を構成する際のアクセント変化は **表-1** に示すように後接単語のアクセント属性によって決定され，文節どうしが結合する際のアクセント変化は先行文節のアクセント型によって決定される[8)~10)]．また同時に，表面的には単純にこれらのアクセント規則が適用できない次のような例もみられ，アクセントの決定には語彙属性，語構成，単語の複合度といったものが深く反映されていることがわかる．

（1） 同じ接辞が違うアクセントを生ずるもの

にほ┐ん＋じん→にほんじ┐ん(日本人)

(アメリカ＋じん→アメリカ┐じん(アメリカ人))

りょ┐うり＋や→りょうり┐や(料理屋)

(おも┐ちゃ＋や→おもちゃや(玩具屋))

（2） 構造をもつ複合語でアクセント規則が再帰的に適用できないもの

び┐じゅつ＋せんもんか→びじゅつせんもんか(美術専門家)

で┐んぱ＋ぼうえんきょう→でんぱぼうえんきょう（電波望遠鏡）

（3） 単語の連接であるが文節アクセント結合となっているもの

ほ┐んじつ＋かいてん→ほ┐んじつかいてん（本日開店）

ふじん＋どうはん→ふじんどうはん（夫人同伴）

実用上は，これらに関与する単語に特殊なアクセント属性を与えるなり，複合語として辞書に記載するなどして大半は救済可能である．アクセントに関連する問題としては，このような例への対処のほか，長い複合語にみられるアクセント句分割，文節中の副次アクセント生起規則といった問題がある[11)]．これらの問題は，音声合成のために必要な処理であると同時に，形態素とアクセントの本質的な関係の分析[12)]といった音声言語学的にも重要な研究課題である．次章で述べる句の構造とイントネーションの関係とあわせて，音声言語学者と計算機科学者の取組みにより本質的な解決が期待される．

3. イントネーション制御と言語情報

自然な日本語音声を合成するためには，韻律の制御が必要であるが，中でもとりわけイントネーションの制御が重要である．イントネーションは，音響的な物理的変数としては，声帯の振動数（基本周波数）によって担われている．**図-2** の例に見られるように，文を発話した場合の基本周波数の時間変化パタンは，1～2文節程度が一まとまりとなって形成する局所的な起伏（アクセント句）と，発話の開始から時間とともに緩やかに下降する大局的な変化によって表現することができる[13)~15)]．前節で述べたアクセント変化に対応する音響的特徴は，基本周波数の局所的な起伏の変化としてあらわれ，依存関係によって支配されている[13)~15)]．図-2 の例にもみられるように，多くの場合先行する句が後続する句を直接修飾する場合

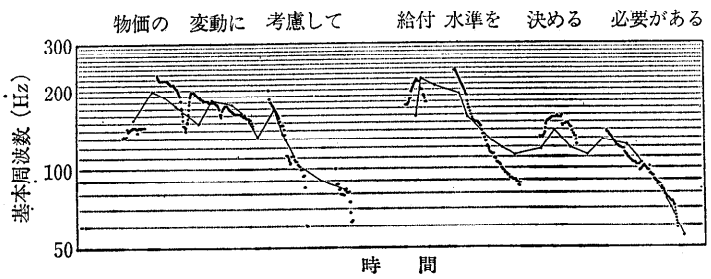


図-2　イントネーションを表す基本周波数の観測パタンと入力言語情報と制御規則によって生成されたパタン（折れ線）

（左枝分かれ構文）は大局的な下降特性が保たれるが，直接修飾しない場合（右枝分かれ構文）は下降せず，いわゆる基本周波数の「立て直し」現象が生ずる．

この文節間の依存関係，いわゆる「係り受け」関係は高度の文章解析が必要である．係り受け関係は，品詞の連接関係や助詞の細分類を用いることによってある程度の推定が可能であり，実際のシステムではヒューリスティクスによる詳細な規則[16]や統計的な立て直し特性の分析結果を用いて実際の問題に対処している[17),18]．よく知られるように，これらの制御に入力として用いる「係り受け」関係は，句構造を文法によって明確に規定してゆく上で必ずしも扱いやすいものではない．文法によって規定される構造と韻律制御を反映した構造との関係の解明が課題である．明確な文法に基づく構造と実際の「立て直し」現象の解明を図る一つの方法としては，構造表記が与えられた多数のデータから自動的に文法を獲得する帰納学習[19]の利用などが有用と思われる．

4. タイミング制御と言語情報

分かりやすく自然な音声を合成するためには，基本周波数とあわせて，テンポ・リズム・ポーズといった音声の時間特性の制御が重要である．日本語音声では，ほぼ仮名一文字に対応するモーラ(拍)を単位としたタイミング制御が図られているが，モーラに対応する音響区分長（音韻長）は決して音声中で一定ではないことが確認されている[20)~23]．**表-2** に示すように，音韻長に影響を及ぼす要因としては種々のものが知られており，モーラ構造やモーラ数，位置といったものの影響が無視できない．また，音声が伝える「ことば」としての役割に依存する要因の存在も統計的に明らかとなっている．

表-2　音韻継続時間長に影響を与える要因

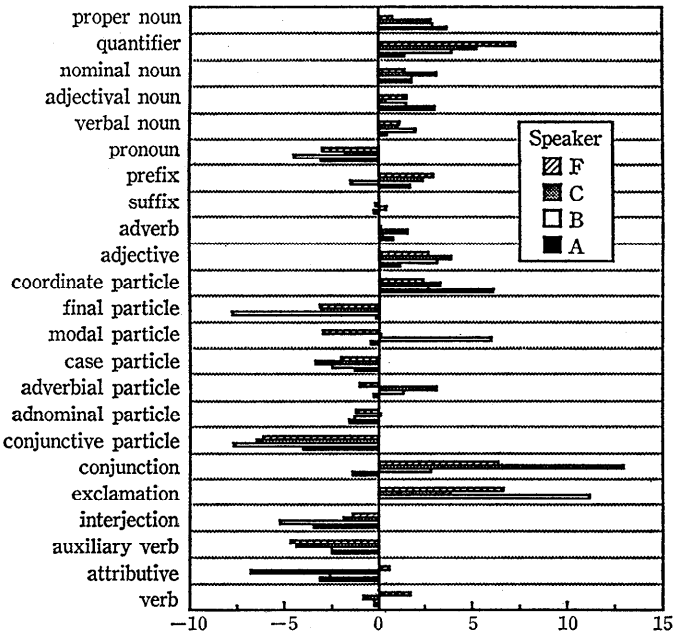
影響範囲	観測される音響的特徴	影響要因
当該音韻	固有平均長 伸縮傾向の相違	調音上の制約
近傍音韻 モーラ	隣接音韻間の時間長補償 長短リズム	モーラ・ タイミング
単語	内容語伸張・機能語短縮	単語の重要度
発話区分 頭・末尾	句・呼気段落末伸長 区分頭短縮	発話区分 境界の明示
発話区分 全体	句・呼気段落内モーラ数 増加にともなう音韻長短縮	発話区分内 テンポ
文全体	発話区間全体の伸縮	発話テンポ

図-3 に，各音韻が属す品詞種類による平均的な音韻長伸縮傾向を示す[23]．この図にみられるように，伝達内容として重要な情報となることが多い名詞などの自立語は伸び，助詞，助動詞などの付属語は短くなる．自立語の中でも，数量詞，固有名詞のように特に重要な情報を担うことが多いものは伸長の度合いが大きい．また，付属語の中でも，特別の文章構造や話題の焦点を示すのに用いられる並列助詞，副助詞などは逆に伸長する．同様な傾向は，音声振幅の制御特性にもあることが知られている[24]．以上のように，情報伝達のために必要な役割に応じた調節がタイミングや振幅の制御にも反映されている．

一方，このような明確な伸縮傾向も，その平均的な絶対値は小さく，英語などにみられるほどの大きな変化ではない．これから分かるように，日本語の場合，モーラを単位としたタイミング制御が大きな制御原理となって働き，他の情報伝達のために用いられる制御の自由度は制限される．この傾向は単に朗読音声だけではなく，模擬会話音声でも観測されている[25]．

5. 音声合成に必要な自然言語処理の展望

以上，音声合成に必要な自然言語処理技術と音声合成のための制御における言語情報の使われ方について現状を紹介した．情報伝達に最適な自然な音声を出力するためには，伝達内容を理解し，目的にあった文表現で，適切な話し方の音声が望ましいが，この実現には多くの難しい問題の解決



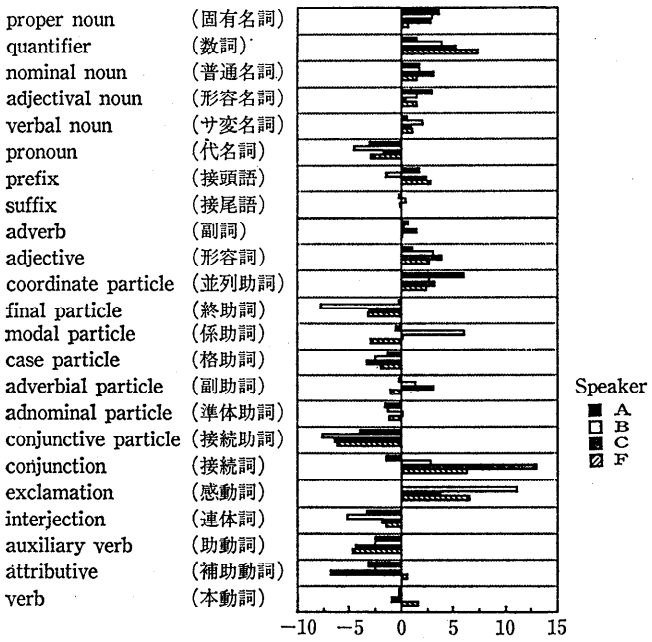


図-3 品詞の違いによる平均的音韻長伸縮傾向

を必要とする．特に，所望の音声出力の具体的な記述と，それを可能とするために必要な入力情報を明確に規定することは難しい．これまでの章でも述べたように，われわれが用いている文章テキストは音声合成にとって必ずしも望ましい入力形式でなく，それに付加する情報もしくは代わる情報表現を見出すことが音声合成の重要な研究課題であると思われる．

このような観点から，音声合成に必要な自然言語処理を考え直してみると，従来からのテキストからの音声合成にとらわれることなく，伝えたい内容を的確に相手に伝達するための音声言語表現の問題として広くとらえることが重要と思われる．本文中でも述べたように，この問題は次のような基本的な研究課題を含む．

（1） 人間が入力しやすく計算機が扱いやすい音声言語情報表現の決定

（2） 文音声の発話に関与する言語情報の特定とそれを用いた制御モデル化

（3） 音声による情報伝達に適した発話文の生成

これらの課題は，自然な合成音声出力のための表記規定，補助記号の設定といった実用を意識した卑近な問題から，人間が発話時に行っている言語処理のメカニズム解明，発話文法の科学的モデル化といった人間の言語行動そのものの探究に繋がる深遠な問題へと広がっている．さらにこれらは，ワードプロセッサの普及をはじめとする情報化時代に即応してゆくための入力標準化といった問題が象徴するように，単に工学の分野だけの問題にとどまらず，工学的実現のための社会的な施策（教育を含めた）と深く結びついている．すなわち，漢字表記の制限，送り仮名，句読点の使用法といった古くから国語学，教育学の分野で論じられてきた問題をはじめ，伝達する相手に分かりやすい文章の書き方，話し言葉としての作文技法が，必要とする自然言語処理の問題点を左右する．現実にある問題を受身的に解決していくだけでなく，積極的な代替案の呈示が必要である．このためには，人間同士の情報交換をより円滑に行うた

めのプロトコル作成の観点から言語表現をとらえなおし，科学的な考え方に基づいた言語使用法を提案し，議論を重ねていくことが急務と思われる．

参考文献

1) 小暮，嵯峨山，佐藤：音声合成のための日本語テキスト解析，音響学会春季講演論文集 1-5-18, pp. 79-80 (1982).
2) Sagisaka, Y. and Sato, H.: Word Identification Method for Japanese Text-To-Speech Conversion System, Proc. ICASSP, pp. 2411-2414 (1986).
3) 宮崎正弘：係り受け解析を用いた複合語の自動分割，情報処理学会論文誌，Vol. 25, No. 6, pp. 970-979 (June. 1984).
4) 宮崎正弘：単語間の意味的結合関係を用いた複合語アクセント句の自動抽出，電子通信学会論文誌，Vol. J68-D, No. 1, pp. 25-32 (1985).
5) 武田，藤崎：統計的手法による漢字複合語の自動分割，情報処理学会論文誌，Vol. 28, No. 9, pp. 952-961 (Sep. 1987).
6) 佐藤，匂坂，小暮，嵯峨山：日本語テキストからの音声合成，研究実用化報告，Vol. 32, No. 11, pp. 2243-2252 (1983).
7) 佐藤大和：連濁の分析と規則化の検討，音響学会秋季講演論文集 1-2-10, pp. 61-62 (1983).
8) 日本放送協会編：日本語アクセント辞典 (1966).
9) McCawley, J. D.: The Phonological Component of a Grammar of Japanese, Mouton (1968).
10) 匂坂，佐藤：日本語単語連鎖のアクセント規則，電子通信学会論文誌，Vol. J66-D, No. 7, pp. 849-856 (1983).
11) Sagisaka, Y. and Sato, H.: Some Accentual Characteristics in Japanese Phrases and Long Compounds, J. Acoust. Soc. Jpn., Vol. E 7, No. 1, pp. 65-74 (1986).
12) 定延利之：現代日本語東京方言における合成的字音語のアクセントと字数について，NEBULAE Vol. 15, pp. 140-157 (1991).
13) 箱田，佐藤：文音声合成における音調規則，電子通信学会論文誌，Vol. J63-D, No. 9, pp. 715-722 (1980).
14) Fujisaki, H. and Kawai, H.: Realization of Linguistic Information in the Voice Fundamental Frequency Contour, Proc. ICASSP, pp. 663-666 (1986).
15) Kubozono, H.: The Organization of Japanese Prosody, Ph. D. thesis (Univ. of Edinburgh), (1987).
16) 広瀬，藤崎，河井，山口：基本周波数パターン生成過程モデルに基づく文章音声の合成，電子情報通信学会論文誌，Vol. J72-A, No. 1, pp. 32-40 (1989).
17) 箱田，中嶌，広川：文章音声の音調結合型導出規則の検討，信学技報 SP 89-5, pp. 33-38 (1989).
18) 海木，匂坂：局所的句構造に基づく F_0 制御，電子情報通信学会音声研究会資料 SP-92-6, pp. 41-46 (1992).
19) Pereira, F. and Shabes, Y.: Inside-Outside Reestimation from Partially Bracketed Corpora, Proc. ACL, pp. 128-135 (1992).
20) 樋口：日本語連続音声における単音の持続時間に関する研究，学位論文（東京大学）(1981).
21) 匂坂，東倉：規則による音声合成のための音韻時間長制御，電子情報通信学会論文誌，Vol. J67-A, No. 7, pp. 629-636 (1984).
22) Takeda, K., Sagisaka, Y. and Kuwabara, H.: On Sentence-Level Factors Governing Segmental Duration in Japanese, JASA, Vol. 86, No. 6, pp. 2081-2087 (1989).
23) 海木，武田，匂坂：言語情報を利用した母音継続時間長の制御，電子情報通信学会論文誌，Vol. J75-A, No. 3, pp. 467-473 (1992).
24) 三村，海木，匂坂：統計的手法を用いた音声パワーの分析と制御，音響学会論文誌，Vol. 49, No. 4, pp. 253-259 (1993).
25) Sagisaka, Y. and Kaiki, N.: Prosody Control for Spontaneous Speech Synthesis, Proc. XII ICPhS, Vol. 3, pp. 506-509 (1991).

（平成5年5月6日受付）

匂坂　芳典

昭和 48 年早稲田大学理工学部物理学科卒業．昭和 50 年同大学院修士課程修了．同年日本電信電話公社（現，NTT）武蔵野電気通信研究所入社．昭和 61 年より国際電気通信基礎技術研究所（ATR）に出向．現在，ATR 音声翻訳通信研究所，第一・第二研究室長．工学博士，音声合成を中心とした，音声情報処理，言語情報処理の研究に従事．「Speech perception, production and linguistic structure」（共編 オーム社），電子情報通信学会，日本音響学会，IEEE，米国音響学会各会員．