解 説 音声処理技術とその応用

3. 音声合成技術

Speech Synthesis Technologies by Keikichi HIROSE (Department of Information and Communication Engineering, School of Engineering, University of Tokyo).

広瀬 啓吉[1]

1 東京大学工学系研究科電子情報工学専攻

1. はじめに

人間の発声機構を解明し，それを機械的に実現しようとすることへの興味は大変に古く，声道に見立てたゴム管を手で変形して母音を合成しようとする試みが，今から200年ほど前にすでに行われていた．このように理学的興味の対象であった音声合成が，工学的な対象となったのは，半導体集積回路技術が進歩をとげたここ20～30年ぐらいのことである．とくに，単語の綴りと発音を音声を提示しながら教育する玩具が米国で発表された後，音声合成への関心が急速に高まった．これは，決まった語句の音声を出力するだけのものであったが，任意のテキストを音声化するテキスト音声合成の試みも1960年代の後半から始まり，1970年代の後半には実用的な英語音声合成システムが開発された．その後，日本語や各言語についてテキスト音声合成システムが開発され，現在では，パーソナルコンピュータのソフトウェアとして一般的になっている．さらに，音声を情報伝達媒体としたマンマシンインタフェースの要素技術として，文章を生成しつつ音声合成を行う概念からの音声合成が重要課題となっている．ここでは，テキストあるいは概念からの音声合成について概説した後，主に音響面での処理を中心に研究の現況と動向を述べる．さらに，音声合成に関連した最近の話題として，声質変換，音声モーフィングについても言及する．

2. テキスト音声合成と概念音声合成

漢字かなまじり文章で書かれたテキストを入力とした場合，音声波形を生成する処理の前に，テキストがどのような単語から構成されているか，主部と述部の境界はどこにあるか，といった情報を抽出する言語レベルでの処理，さらにこれらの言語情報と音声の音響的特徴とを関連づける操作が不可欠である．一般に，テキストから音声を合成するためには，**図-1**のように言語処理，音韻処理，音響処理の各段階の処理が必要となる．具体的な処理内容は個々の合成システムにより細かい違いがあり，最近の統計的手法による音声合成では段階の区分が，明確でない場合もあるが，その概略は以下のようにまとめられる[1]．

言語処理：形態素解析によってテキストを単語単位に区分し，品詞情報を出力するとともに，統語・意味・談話解析によって統語情報・談話情報などの言語情報を出力する．しかしながら，高次の言語情報を誤りなく抽出することは，現在の技術では困難である上，人間の発話特性の結果，それが必ずしも正確に音声現象に現れるとはかぎらない．そのため，実際の音声合成システムでは隣接語句の関係程度の情報を抽出して利用することが多い．

音韻処理：発音辞書にアクセスしてテキストを実際の発音の表記に変換すると同時に，単語・文節のアクセント型を導出する．次に，言語情報と韻律の関係を記述する規則などを用いて，合成音声の韻律の特徴を生成する．

音響処理：音韻処理で得た発音表記に従って合成単位を蓄積したデータベースにアクセスし，必要なものを取り出し，接続して音声波形を生成する．合成単位は，音響パラメータあるいは波形として蓄積されており，前者の場合は，接続して得たパラメータ時系列で合成回路を制御して音声を合成する．

最近，音声により応答する対話システムの開発が進んでいる．簡単なシステムでは，定形文に応答内容の語句を挿入する録音編集によって応答音

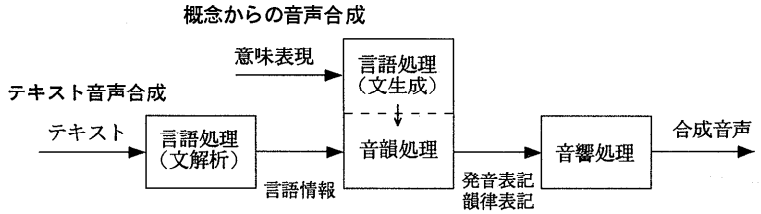


図-1 テキストからの音声合成と概念からの音声合成の処理の流れ

声を生成するが，システムが高度になり，多様な内容の文で応答するためには，システムは，まず，ユーザに伝えたい内容の意味表現を生成し，それを音声化して出力する必要がある．この場合，システムは，図-1のように意味表現から表層文を生成すると同時に音声合成を行う．このような概念からの音声合成では，音韻処理，音響処理の部分はテキスト音声合成と同じであるが，言語処理の部分は大きく異なる．文解析のかわりに文生成のプロセスが必要となるが，その過程で高次の言語情報が誤りなく得られ，これを合成に利用することによって，品質の高い音声を得ることが可能となる．すでに述べたように，従来のテキスト音声合成システムでは，このような情報が得られることを想定していないために，その合成規則をそのまま用いることはできない．新しい合成規則の開発が必要となる．

合成音声の品質向上のためには，もちろん，各段階の処理の高度化が必要であるが，以下では，音声に直接関連した音韻処理以降の段階について，音声の分節的特徴☆と韻律的特徴☆2の2つの側面から解説する．

3. 分節的特徴の合成

音韻処理で中心となるのが単語単位での読みの決定である．漢字には，音読み，訓読みといわれるように一般に複数の読み方がある．形態素解析が正しく行われている時は，多くの場合，発音辞書により簡単に読みを決定できるが，未知語では読みの推定が，連濁，接辞では音素表記の変化への対処が必要となる☆3．さらに，音素記号を実際に発音に対応する読みに変換する際にも，両者が1対1に対応しないという，いわゆる異音化の問題がある☆4．音韻環境を考慮した処理が行われる．

音韻処理で得られた読みから対応する音声の特徴パラメータ時系列を生成する手法としては，発話速度や声質の柔軟な制御が可能，必要なメモリ量が小さいなどの点から，音声の生成過程に従った規則合成が理想的である．しかしながら，このような生成過程に関する知見は，母音連鎖についてはある程度得られているものの，子音が含まれた連続音声に関しては未知の点が多い．幸い，分節的特徴は音素や音節程度の小さな単位で取り扱うことが可能であり，したがって，現状では，自然音声の分析結果をもとに作成した音節などの特徴パラメータパターン(あるいは直接に音声波形)をデータベースに蓄積しておき，合成に際しては，読みに従って必要なものを取り出して接続することが一般的に行われている．

合成単位の接続は合成音声の品質の低下の大きな要因となる．この観点からは，なるべく大きな単位での合成が望ましいが，反面，蓄積単位数の著しい増加を招く．蓄積単位の作成は，自然音声を切り出すことによって行うが，その発声単位は，調音結合の影響を考慮した場合，母音－子音－母音の連鎖以上が必要となる．話者の発声のしやすさからは単語程度が望ましい．単語音声を利用する場合，合成対象と長い区間で一致する音素系列が単語音声データ中に得られる可能性があり，そのような音素系列を利用する方が画一的な単位を利用するよりも効率的であり，音質もよい合成音声が得られると考えられる．このような観点から，複合音声単位の合成手法が提案されている[2]．

3.1 合成方式

音声波形を合成する方式としては，音声波形の信号処理を行う波形編集方式・分析合成方式，音声の生成過程に着目したターミナルアナログ方式・声道アナログ方式がある．以下，それらについて概説するが，後2者については，後述するよ

☆ 音素などの単語や文を構成する個々の音の特徴．主に声道伝達特性としてスペクトルの包絡により表される．

☆2 抑揚やリズムを表す音響的特徴．基本周波数，持続時間，パワーなど，主に，音源に関する特徴である．

☆3 連濁の例：めざまし＋とけい→めざましどけい，接辞の例：一＋本→いっぽん．

☆4 1つの音素が音韻環境で異なった音として発音される現象．たとえば，neNbaNgaNの3つのNではそれぞれ音の特徴が異なる．またgは鼻音化する．

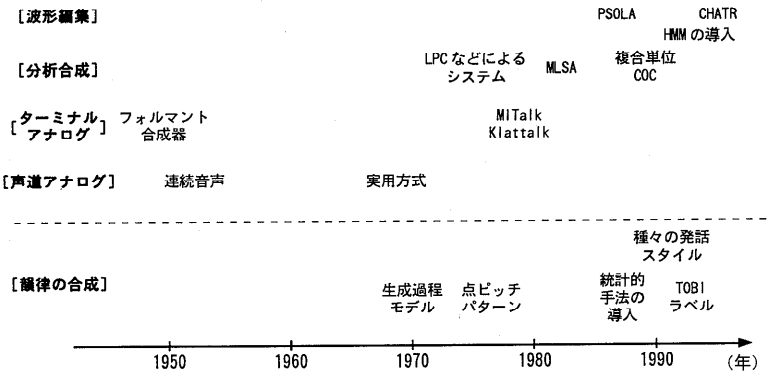


図-2 音声合成手法の歴史的な流れ

うに，統計的手法の導入により，現在の合成研究の主流となっている．図-2に合成手法の流れを簡単に記した．

波形編集方式：自然音声波形から，合成単位の音声波形を，前後の音素環境，韻律的情報などの情報とともに切り出し，波形辞書として蓄積する．合成時には，音韻環境がテキストの音韻処理結果と最も合致する波形を選択して接続する．波形そのものを用いるため，分析・合成処理にともなう品質の劣化がないが，多量の音声データを効率よく取り扱うことが重要である．

自然音声波形そのものを用いるために個々の合成単位の品質は高いが，韻律の制御に難点があり，従来は，駅の案内や時報の案内のように，数種の単語音声を定形文の中に埋め込んで出力するといった限られた用途に用いられていた．これに対し，Time Domain Pitch Synchronous Overlap and Add 法(TD-PSOLA 法)が，diphone(音素の中心から次の音素の中心までの単位)を用いた合成システム(フランス語)として報告され[3]，高い品質が得られたことから，合成に広く用いられるようになった．この方法は，まず素材の波形にその基本周期☆5に従ってマーク(ピッチマーク)をつけ，それを基準として，波形を基本周期の2倍程度のハニング窓で切り出し，合成音声の基本周期間隔で重ね合わせを行うものである．合成に用いる波形の基本周波数を低くする場合は，切り出された波形があまるので適宜に間引きし，逆に，高くする場合には適宜に繰り返して用いる．ピッチ同期処理であるが，基本周期の抽出誤りによる音質の劣化は小さい．ピッチマークは波形の1周期中での振幅最大点に設定することが多いが，調波構造のため安定して抽出することが困難で，合成音声にゴロゴロ感を与える結果となっている．この有効な解決策として，ウエーブレット変換を用いて声門閉鎖点を高精度に抽出し，これを波形重ね合せの基準点とする手法などが提案されている[4]．

PSOLA 法として，波形操作を分析合成により周波数領域で行う FD-PSOLA 法がある．後述の分析合成方式のように分析パラメータレベルで接続してから合成する場合には，そのパラメータの連続性が問題となり，一般には線形予測分析など，音声の生成モデルを前提とした分析を行う必要があるが，FD-PSOLA 法ではそのようなことはなく，スペクトル包絡☆6を現象的に再現する分析手法の利用が考えられる．たとえば，最近，スペクトルを時間軸と周波数軸の2次元上に広がった曲面と考え，これに平滑化操作を施すことにより，スペクトル包絡を求める STRAIGHT と呼ばれる分析合成方式が提案され，基本周波数の変化に対する音質の劣化が小さいことが報告されている[5]．これを利用して合成単位ごとに波形操作を行い，その後，接続することが考えられる[6]．

分析合成方式：線形予測法やケプストラム法などによって音声を分析し，スペクトル包絡特性と音源特性に分離する．これを音節程度の単位で蓄積し，必要に応じて取り出して接続することにより，連続音声の制御パラメータ時系列を得る．音源特性が分離されているために，基本周波数制御

☆5 声帯振動の周期で波形上では繰り返し周期として観測される．

☆6 有声音声のスペクトルは声帯振動に対応する線スペクトル構造をもつ．簡単にはこのようなスペクトルのピークを結んだ包絡のことで，物理的には声道伝達特性に対応．

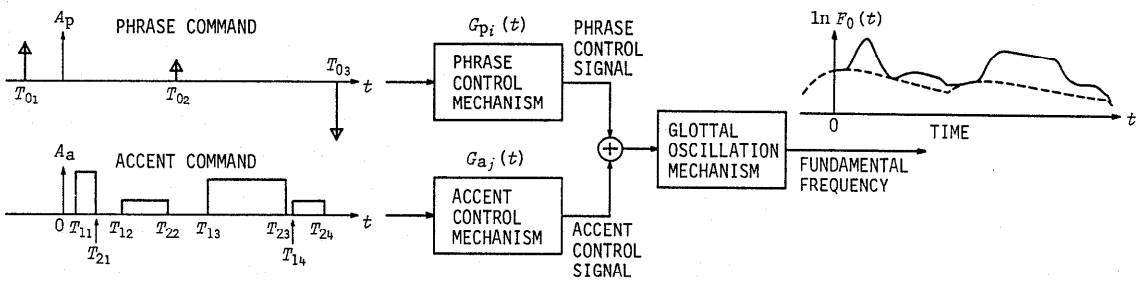


図-3 基本周波数パターン生成過程の重畳型モデル

などが容易である．

ターミナルアナログ方式：音声の生成過程を周波数特性レベルで模擬することによって合成する．声道の伝達特性は複数個の極および零点の特性を掛け合わせたものとして表されるが，これを，極に対応する共振回路，零点に対応する反共振回路を組み合わせることによって模擬し，音源波で励振する．母音の場合の共振周波数はフォルマント周波数として与えられるなど，音声の物理的特徴との対応が直接的であり，規則による合成に適している．蓄積パターンの接続による場合でも，百数十個の音節パターンを用意することで，比較的品質の高い音声合成が可能である[7]．

声道アナログ方式：声道内の音波の伝搬特性にまでさかのぼって模擬する音声合成方式である．調音との対応を，ターミナルアナログ方式よりもより直接的にとることが可能であり，言語情報との対応もつけやすいと考えられるが，その反面，規則の導出に必要な，正確な生理的データが得にくいという欠点がある．

4. 韻律的特徴の合成

分節的特徴は基本的にはテキストに示されているのに対し，韻律的特徴は，通常，句読点以外には明記されていない．したがって，何らかの手段によって，それを導き出す必要があるが，そのためには辞書引きによる単語のアクセント型の決定だけでは不足であり，文の統語構造や文章の談話構造の利用が不可欠である．

連続的な文音声を発声した場合，その韻律的特徴は構成単語のそれとは大きく異なることが多い．これは，たとえば，連続音声中では，単語の基本周波数パターンのアクセント成分が，隣接韻律語のアクセント型，文の統語構造，文章の談話構造の影響を受けて融合，増大，減少，消滅などの複雑な変化を示すためである．また，文のイントネーションに対応する基本周波数パターンのフレーズ成分も統語構造と深くかかわっている[8]．このような言語情報と韻律的特徴との関係を規則として表現するためには，両者を明確に結びつけ得るモデルが必要である．図-3のように，基本周波数パターンをフレーズ成分とアクセント成分との対数周波数軸上での和として表現する生成過程の重畳モデルが提案され[9]，これを用いてアクセント・イントネーションの観点から優れた品質の合成音声が得られている[7]．

4.1 統計的手法の利用

従来，言語情報と韻律的特徴とを関係づける規則は，音声の分析結果をもとに，人間が発見的に構築していたが，最近では，数量化手法，さらには音声認識で一般的な隠れマルコフモデル(HMM)やニューラルネットワーク(NN)を用いて音声データから自動的に規則を生成することが行われている．とくに，音素の持続時間に関しては，種々の要因が関係していたために発見的手法では定式化が困難であり，数量化手法が有効である．持続時間に対する各制御要因(音韻環境，文内位置，モーラ数，品詞など)の関与の度合を数量化I類によって調べ，要因から音素持続時間を与える線形回帰モデルが作成されている．さらに，進んだモデル化として，回帰木により要因を適宜組み合わせることが行われている[10]．

数量化手法を，基本周波数パターンに利用するためには，何を数量化の対象とするかの工夫が必要である．たとえば，前述の基本周波数パターンの重畳モデルでは，フレーズ成分，アクセント成分が，それぞれ指令の大きさとして与えられるので，それを用いることが考えられる．ここで，数

量化に際して，多量の分析済みデータが必要なことが大きな問題となる．重畳モデルは声帯の物理的特性を考慮した優れたモデルであるが，観測される基本周波数パターンからモデルのパラメータを解析的に抽出することができず，逐次近似法によっている．このため，人間が介在した分析が行われ，データベースの整備が進んでいない．一方，英語の韻律を記述する指標として単語境界情報，抑揚情報などを表記する Tone and Break Indices (ToBI)が開発されているが，最近，これを日本語に応用し，韻律データベースを整備することがATRで進められている．これも人手をまったく介さないというものではないが，モデル分析よりもデータベースの規模の拡大は容易であり，これを基に韻律の合成規則を作成し，得られた指標を用いて，重畳モデルにより基本周波数パターンを合成することが考えられる[11]．

HMMを用いて基本周波数パターンを生成することが行われている．重畳モデルのような生成モデルを必要としないため，多量のデータを用意することが容易という利点がある．語句の基本周波数パターンの特徴は当該語句，先行語句の言語的属性などにより異なるが，これをHMMのノード・アークの分割よって表現する手法が提案されている[12]．言語解析を明に行わないという特長を有する．NNを利用した場合も同様にモデルを必要としないという利点があり，言語処理にも有効である．

4.2 種々の発話スタイルの音声合成

現在のテキスト音声合成では，朗読調の高品質音声を合成することに研究の主眼が置かれているが，将来的に音声対話システムに合成音声を利用することを考えると，対話調をはじめとしてさまざまなスタイルの音声を合成する技術の開発が求められる．対話音声や感情音声の韻律的特徴を調べ，規則化することが進められているが，韻律的特徴の変化のレンジが大きく，また，関係する要因も多くなるために，必ずしも容易な課題ではない．分析が進んでいる朗読音声と比較することによって，その特徴を明確に捉えることができる．

5. 音声データベースを用いた合成

音声合成は，音声生成過程を再現するという物理的な側面の研究からスタートした．このため，当初は，ターミナルアナログ方式，声道アナログ方式による開発が中心であった．このような規則音声合成技術は，ごく少数の蓄積パターンから多様な音声を合成し得るという利点があって，大変期待され，現在も大学を中心として引き続き研究が行われているが，音声生成過程の正確な把握が困難なこともあって，今のところ，思ったような品質の音声が得られていない．計算機のメモリ容量が拡大するにつれ，これにかわって，多量の音声データベースから合成環境にあった合成単位を選択し，波形編集方式や分析合成方式を用いて接続・合成することが盛んに行われるようになった．さらに，すでに音声認識で導入され，成果があがっている統計的手法に目が向けられるようになった．このような音声データベースと統計的手法に基づく最近の合成手法について，以下に研究の流れを説明する．(韻律については**4.1**節を参照)

多量の音声データベースを用いた場合，そこから合成単位を適切に選択し得るか否かが，音質の良否に大きく影響する．連続音声中の音素の特徴が隣接音素の特徴によって変形することは，調音結合現象としてよく知られている．このため，合成の際の音韻環境になるべく一致した単位を選択することが望ましく，不一致の程度を示す歪みを単位選択のコストとする．また，隣接単位間のスペクトルの不連続の程度も音質に影響し，これもコストに加える．コスト最小となるように単位を選択するが，これを効率的に行う種々の工夫が報告されている[10]．

調音結合に対処するためには，少なくとも当該音素に前後の音素を考慮したTri-Phoneをすべて用意して合成することが望ましいが，これは，日本語では15000にも及び，現実的ではない．そのため，適宜，まとめて単位数の削減を図る必要がある．これを効率的に行うために，当該音素のデータベースを音韻環境でクラスタリングするCOC (Context Oriented Clustering)などが提案されている[13]．

合成音声の品質が劣化する大きな原因は，音韻処理結果から得られる韻律パターンを再現するように合成単位を変形することにある．このような観点から，従来とは異なり，韻律の観点でバラエティーに富む文音声データから合成単位を選択す

ることにより，波形の加工を基本的に行わないCHATRと呼ばれる合成方式が，最近開発された[14]．非常に品質の高い合成音声が得られているが，韻律の面からの問題点の指摘もある．

合成に用いる音声データには音韻情報があらかじめラベリングされている必要がある．当初は，これは人手で行われていたが，非常に労力のかかる作業である．発話内容が表記されていれば，HMMを用いて音声のセグメンテーションとラベリングを自動的に行うことが可能である[15),16]．まず，文脈独立の音素HMMを用いて，モデルの学習とセグメンテーションを交互に行って，HMMとセグメンテーションを最適化した後，次に文脈依存の音素HMM（Tri-Phone Model）の学習を行う．この際，前述したように，クラスタリング手法によりモデル数の削減を図る．合成に際しては，音素HMMあるいはその各状態ごとに平均値に近い特徴の波形データを選択して用意しておき，PSOLA法により接続して合成する．人間がラベリングした場合と比較して遜色のない品質の音声が得られている[17]．波形の接続によらず，音素HMMを接続してパラメータ時系列を生成し，音声合成をすることも行われている[18]．

6．声質変換と音声モーフィング

波形編集方式を用いることにより，ある特定の話者らしい音声を合成することが簡単に行えるようになった．しかしながら新しい話者に対しては，再度音声データを取り直す必要がある．そこで，このような音声データの取り直しを行わずに，新しい話者の少量の音声データをもとに，声質の変換を行う声質変換技術が研究されている．また，最近では，画像のモーフィングにヒントを得て，話者Aから話者Bへの変換を連続的に行う音声モーフィングという考え方が登場してきた[19]．声質を決定する要因としては，声道特性に起因するものと，音源特性に起因するものがあり，両者に着目して声質変換を行う必要があるが，ここでは前者に関してのみ述べる．話者Aと話者Bの間でVQコードブック☆7の対応（マッピング）をとり話者の変換（適応）を行うことは音声認識で従来から行われていたが，これを利用して話者A，B間の声質変換が可能である[20]．音声モーフィングは中間的な特徴の音声を生成する必要があり，コードブックマッピングは利用しにくく，たとえば，スペクトル包絡について，話者A，B間の対応をとり，連続的に変化させることで実現できる．

声質変換にしても音声モーフィングにしても，音声の特徴を大きく変形するために，現状では得られる音声の品質に問題が多い．今後の研究成果が待たれる点である．

☆7 ある話者の音声の特徴空間をベクトル量子化し，各ベクトルに符号を与えたもの．

7．合成音声の品質

以上では，単に高い品質の合成音声というような漠然とした記述をしてきた．これは，おおざっぱには人間の発声に近い音声ということを意味しているが，詳しくみると，聞き手にとって自然な音声ということと，分かりやすい音声ということは多少異なる．合成音声が分かりやすいか否かの指標については，高能率伝送での音声の評価指標として確立している音節明瞭度，単語了解度が利用できる．これは，合成音声を聞かせ，その音節あるいは単語が誤りなく聞こえた割合を求めるもので，客観的な評価が得られるという優れた点があるが，文などのより大きな単位での評価手法は確立していない．抑揚の観点からの評価も加味した手法の開発が必要である．文や文章を対象とした場合，音節明瞭度，単語了解度が悪くても，文脈により聞き取れることが多く，聞き手にほかの作業をさせて負荷をかけた状態で聴取実験するなどの工夫が必要である．一方，合成音声の自然さについては，客観的な評価が困難であり，聞き手の好みも加味された主観的な評価とならざるをえない．合成音声の利用といった観点からは，長時間，聞いた場合，負担となるか否かといった観点からの評価も重要である．音声合成技術の進展に対応して，合成音声の評価手法の確立が，今後の重要な研究課題の1つとなっている．

8．おわりに

最近の音声合成研究は，多量の音声データベースを統計的手法で処理するようになっており，この観点からは，音声認識と似通ってきている．これは，音声の物理的現象の正確な把握がなかなか進展しないことにも起因しているといえるが，新しい音声分析法の開発とそれに基づく音声生成モ

デルの構築といった地道な努力も合成音声の品質向上に重要である．

音声合成の品質というと，従来は，まず，その分節的特徴の良し悪しが問題になり，それに比較して，韻律的特徴は軽視される傾向にあった．文字言語にない音声言語の大きな特徴は韻律にあり，この観点から，今後，韻律の合成の重要性が増すと考えられる．

現在，テキスト音声合成は各所で開発され，そのうちのいくつかは，インターネットにより聞くことが可能である．とくに，米国を中心とした海外の音声合成については整理が進んでおり，

http://www.speech.cs.cmu.edu/や

http://www.itl.atr.co.jp/

などから/com.speech/Section5/Q5.4.html へリンクをたどると音声を聞くことができる．日本の合成音声については，筑波大学の板橋秀一教授が中心となってCDにまとめたものがあるが，残念ながら現在，在庫がないとのことである．なお，このような情報を収集するにあたり，音声研究者のメールを利用して多くの方に問い合わせをし，いくつか合成音声を聞くことができるサイトをご紹介いただいたが，最近の音声合成ということで関係が強いもののみを参考文献の欄に記し，そのほかはあえてここでは割愛した．情報をお寄せいただいた方々に深謝する．

最後に，音声関連の参考書としては，数多くの本が出版されているが，その中から参考文献に1つをあげておくので用語などが分からない場合の参考にされたい[21]．

参考文献

1) 広瀬啓吉：音声の出力に関する研究の現状と将来，日本音響学会誌，Vol.52，No.11，pp.857-861 (1996).

2) Sagisaka, Y., Kaiki, N., Iwahashi, N. and Mimura, K.: ATR ν-Talk Speech Synthesis System, Proc. ICSLP 92, pp.483-486 (1992).

3) Moulines, E. and Charpentier, F.: Pitch Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis Using Diphones, Speech Communication, Vol.9, pp.453-467 (1990).

4) 阪本正治，斉藤 隆，鈴木和洋，橋本泰秀，小林メイ：波形重畳法を用いた日本語テキスト音声合成システムについて，電子情報通信学会技術研究報告，SP95-6，pp.39-45 (1995).

(http://www.trl.ibm.co.jp/projects/s7260/protalk.htm)

5) Kawahara, H.: Speech Representation and Transformation Using Adaptive Interpolation of Weighted Spectrum: Vocoder Revised, Proc. IEEE ICASSP 97, pp.1303-1306 (1997).

6) 丁 文，藤澤 謙，ニック・キャンベル，樋口宜男：STRAIGHTを用いたCHATRの韻律制御，日本音響学会秋季講演論文集，1-2-8，pp.211-212 (1997).

7) Hirose, K. and Fujisaki, H.: A System for the Synthesis of High-quality Speech from Texts on General Weather Conditions, IEICE Trans. Fundamentals, Vol.E76-A, pp.1971-1980 (1993).

8) 広瀬啓吉，藤崎博也：音声合成とアクセント・イントネーション，電子情報通信学会誌，Vol.70，pp.378-385 (1987).

9) Fujisaki, H. and Hirose, K.: Analysis of Voice Fundamental Frequency Contours for Declarative Sentences of Japanese, J. Acoust. Soc. Japan (E), Vol.5, pp.233-242 (1984).

10) 匂坂芳典：日本語音声合成の最近の進展について，電子情報通信学会技術研究報告，SP94-84，pp.39-45 (1995).

11) 平井俊男，樋口宜男：韻律ラベリング・システムJ-TOBIを用いた基本周波数制御規則の自動抽出，電子情報通信学会技術研究報告，SP97-5，pp.27-32 (1997).

12) Ishikawa, Y. and Ebihara, T.: On the Global F0 Shape Model Using a Transition Network for Japanese Text-to-Speech Systems, Proc. EUROSPEECH 97, pp.2679-2682 (1997).

13) 中嶌信弥，浜田 洋：音韻環境に基づくクラスタリングによる規則合成法，電子情報通信学会論文誌，Vol.J72-D-II，pp.1174-1179 (1989).

14) ニック・キャンベル，アラン・ブラック：CHATR：自然音声波形接続型任意音声合成システム，電子情報通信学会技術研究報告，SP96-7，pp.45-53 (1996). (http://www.itl.atr.co.jp/chatr)

15) Ljolje, A. and Riley, M.: Automatic Segmentation of Speech for TTS, Proc. EUROSPEECH 93, pp.1445-1448 (1993).

16) Donovan, R. and Woodland, P.: Automatic Speech Synthesizer Parameter Estimation Using HMMs, Proc. ICASSP 95, pp.640-643 (1995).

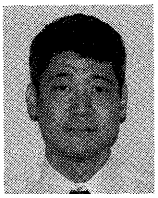
17) Yoram, M. and Hirose, K.: Waveform Concatenation Speech Synthesis Using Phonetic Clustering and Automatic Unit Selection, 日本音響学会秋季講演論文集，2-2-20，pp.263-264 (1997).

18) Tokuda, K., Kobayashi, T. and Imai, S.: Speech Parameter Generation from HMM Using Dynamic Features, Proc. ICASSP 95, pp.660-663 (1995).

19) 阿部匡伸：基本周波数とスペクトルの漸次変形による音声モーフィング，電子情報通信学会技術研究報告，SP96-40，pp.25-32 (1996).

20）阿部匡伸：音声変換処理技術－基本周波数，継続時間，声質に関して－，電子情報通信学会技術研究報告，SP93-137，pp.69-75（1994）．
21）古井貞熙：ディジタル音声処理，東海大学出版会（1985）．

（平成9年9月9日受付）



広瀬　啓吉（正会員）

1972年東京大学工学部電気学科卒業．1977年同大学院博士課程修了．工博．同年東京大学工学部電気工学科講師．1994年同電子工学科教授．1995年同大学院工学系研究科電子情報工学専攻教授．1987年米国MIT客員研究員．音声情報処理の教育・研究に従事．電子情報通信学会，日本音響学会，人工知能学会，IEEE，米国音響学会，ESCAなど各会員．

e-mail:hirose@gavo.t.u-tokyo.ac.jp