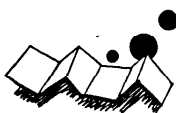


解説

音声認識合成システム†



篠田 英範† 浮田 輝彦†

1. はじめに

人類にとって、鳥のように空を飛ぶことが夢であったように、人間の言葉を聞きかつ話す機械を作ること、また夢であった。近年の計算機技術、人工知能の技術はこの夢をかなえる方向に向かって、着実に歩を進めている。人間の言葉を聞く機械は、音声認識装置として、数百単語を聞き分けるほどに、また、話す技術は、音声合成装置としてかなり流暢に喋るものも商品化されるほどに進歩している。

ここでは、メモリ応用技術の立場から、音声認識・音声合成装置において、どのようなメモリが、どのように利用されているか、また、今後どのようなメモリが望まれるかなどを考える。

ここ数年の傾向を見ていると、メモリの1チップ当りの容量は3年ごとに4倍になっている。今世紀末には64~100 Mビットの容量をもつ超LSIメモリ素子が出現すると予想される。この観点からは、メモリ素子提供者にとって大量にメモリを利用する応用分野の開拓が望まれる所であろう。音声認識・合成装置は入出力装置であり、できるだけコンパクトに作られるべきである。この意味では、大容量のメモリ素子を利用できることは、装置開発者にとって有難いことである。

以下、音声認識・音声合成装置におけるメモリの位置付けについて解説する。音声は、人間同士のコミュニケーションにおいて、最も基本的な手段の一つである。図-1はその音声の生成と知覚の過程を示したものである^{1),2)}。下にたどれば音声合成の過程に、上にたどれば音声認識の過程と考えられる。図の上の部分にかかわるほど複雑な処理が必要となる。合成では、メッセージを構成する単位として、音節、単語、文などの要素のうち、どれを選択して合成処理を行うかに

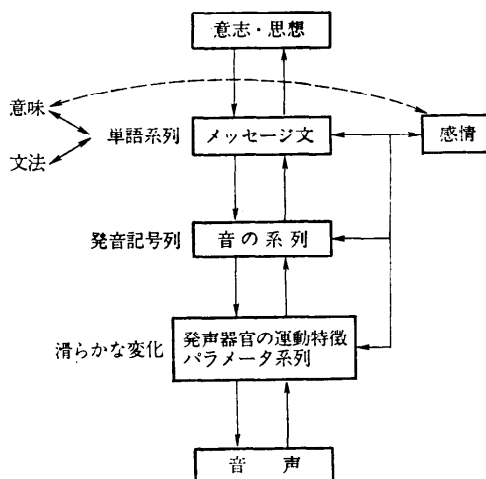


図-1 音声生成と知覚の過程

より、複雑さと装置の柔軟性が決まる。認識では、メッセージ文の構造(発声単位)により複雑さを回避することができる。現在の技術では、音声認識と音声合成との装置構成上の共通基盤はあまりない。そこで、ここでの記述は音声認識装置と音声合成装置とを分けて行うこととする。

2. 音声認識・合成システムとメモリ

この章では音声認識・合成装置におけるメモリの役割について、認識・合成処理の概要とともに解説する。図-2に音声処理の単位と記憶量の関係を示す。処理単位が小さいほど、一般性は増加し、記憶容量も小さくなるが自然性や認識性能の点で課題がある。

2.1 音声認識

音声認識技術は、話者の違い(特定/不特定話者)、発声法の違い(孤立/連続発声)、対象語集数、発声環境などの要因により、種々のレベルに分けられる。

図-3に音声認識処理の概要を示す。単語認識では、認識単位として単語と音素を用いるものに大きく二つに分かれる。単語単位の認識装置では、十数語(不特

† Speech Recognition and Synthesis System by Hidenori SHINODA and Teruhiko UKITA (R & D Center, Toshiba Corp.).

†† (株)東芝総合研究所

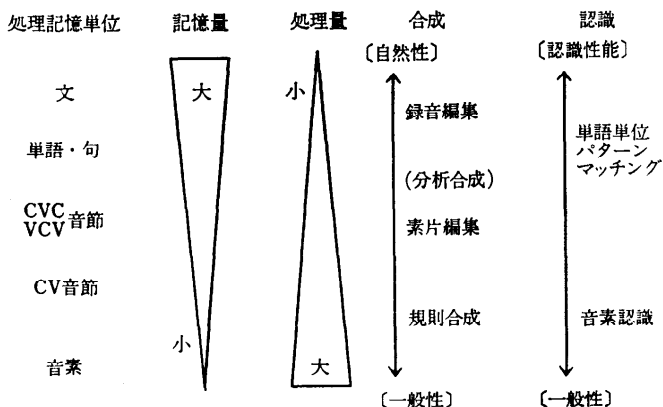


図-2 音声認識合成と処理記憶単位

といった多くのレベルの知識源が総合的に利用される必要があり、膨大なメモリと複雑なシステム構成が必要になる。

2.2 音声合成

音声合成方式は、図-1 のどの位置の情報を記憶し、どのような情報で合成処理を制御するかにより分類できる。

音声波形を記録し合成する方式を、録音編集方式という。波形の記録には、合成された音声波形が元の波形をより正確に再現しかつ記憶容量を小さくする種々の符号化方式が利用されている。

発声器官の運動を制御するパラメータに対応するものを記憶しておき、これにより合成処理を行い音声を生産する方式を、分析合成方式という。この方式では一般に音声生成機構をモデル化した合成フィルタにより音声合成を行い、これを制御するパラメータがメモリに記憶される。記憶容量は録音編集方式より小さい。

メッセージ文の文字列を与えて、対応する音声を作成する方式を規則合成方式という。この場合に記憶される情報は、メッセージ文を解析するための文法、単語辞書、アクセント・イントネーションなどを制御するための韻律規則、発音の仕方を決定する音韻規則、合成フィルタを駆動し音声合成を行うためのパラメータや音声の素片などである。

これらの方式は、それぞれ実用化されている。録音編集方式は音質の良さを生かして駅構内の案内放送などに、分析合成方式は装置のコンパクトさを生かして、時計や玩具に組み込まれている。規則合成方式は、言語情報処理の要素も絡みこれからの技術であるが、DECtalkのように商品化されているものもある³⁾。それぞれの音声合成方式におけるメモリの位置付けを図-4 に示す。

3. 音声認識・合成処理のデータ構造

この章では、音声認識・音声合成システムのデータ構造について解説する。データ構造は方式により異なる。各方式の代表例について述べる。

3.1 音声認識

現在実用化されている単語音声認識技術を中心に、音声認識処理で用いられる知識源のタイプとそのデー

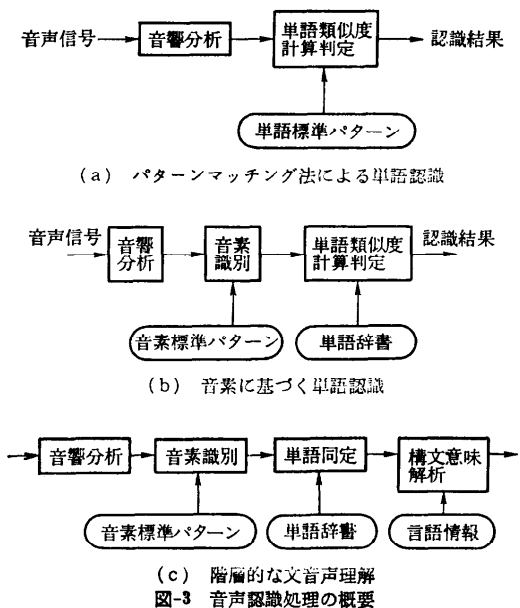


図-3 音声認識処理の概要

定話者)~数百語程度(特定話者)の認識装置及び数十単語の連続単語認識装置が実用化されている。また音声ワードプロセッサとして、離散的に区切って発声した音節を認識することにより文章を入力する装置も商品化されている。使用されるメモリとしては、入力データのバッファ、標準パターンや単語辞書の記憶部、類似度計算のための作業領域が中心になる。

一方本格的な連続音声認識技術は実験段階に止まっている。認識対象が自然な文でその内容を理解しようとすると、単に音響・音声処理のレベルだけでなく、言語情報処理の要素も必要となり、音声理解として研究されている。このときには音響・音声・構文・意味

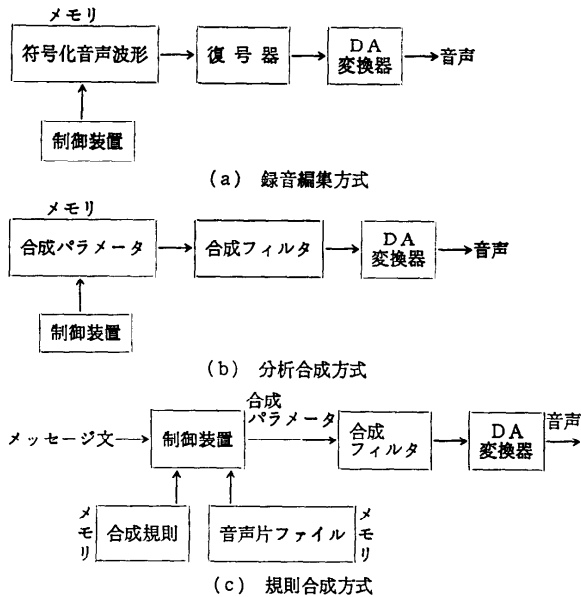


図-4 音声合成システムにおけるメモリの位置付け

タ構造を整理する。

すべての方式に共通の前処理として、入力された音声波から特徴パラメータとして、音声の周波数成分に対応するものが抽出される。フーリエ変換や線形予測分析、また音声帯域を8~30程度の帯域に分割するバンドパスフィルタ(BPF)群などが利用される⁴⁾。これは、8~20ms程度の短区間に対して分析される(フレームと呼ぶ)。これを同程度の時間間隔で、周期的に繰り返すことにより、特徴パラメータの時系列が抽出される。そして、フレームパワーの時間変化などを利用して、音声区間の検出を行った後に、発声内容の判定処理が行われる。

(1) 時間正規化マッチング法による単語認識

図-3(a)のパターンマッチングの手法に基づく方法は、入力を音響分析して得られるスペクトルの時系列を標準パターンとしてあらかじめ用意しておく。入力された発声データとの類似度を計算して、最も類似度が大きいものが発声されたと考える。類似度計算に、動的計画法を利用する時間正規化マッチング方法が開発されている^{5),6)}。標準パターンとは長さ(フレーム数)が異なる入力パターンの発声時間長の変動を非線形な対応付けにより、吸収するものである。

標準パターンは特徴パラメータの時系列そのままの形式で記憶しておく。たとえば、フレーム周期10ms, BPFチャンネル数10チャンネルで、平均単語長が500ms

とすると、1単語当たり、500バイトで表現できる。したがって認識対象語彙が100単語なら、標準パターンの記憶に全部で50kバイト必要となる。また類似度あるいは距離の計算時に部分的な累積値などを記憶するテーブルが必要となるが、多くは必要としない。

また時間正規化マッチングのアルゴリズムは、連続単語音声認識にも拡張されている^{7),8)}。

標準パターンの記憶には、利用者の声の登録作業がたびたび必要となるので、RAMを利用する。この場合、ある個人の標準パターンは不揮発性の外部記憶を利用することにより、保持しておかなければならない。

(2) 統計量に基づいた単語認識

統計的パターン認識法を用いる不特定話者用の単語認識装置が開発されている^{9),10)}。これらの装置では、単語パターンを一つのベクトルとして表現するために、BPFのチャンネルに対応して周波数成分を一定次元と時間軸の成分を一定個数(時間間隔を再構成)を取り出し、ベクトル表現にしている。これにより識別関数や複数の直交ベクトルの形式で多くの統計量をコンパクトに表現する。これらの関数や標準パターンはシステム構築時に1回用意すれば良いので、ROMを用いることができる。

(3) 音素に基づく単語認識

一度入力音声を音素識別した後、単語を認識する方法である。音素識別の段階では完全な認識は望めないで、曖昧さを許した形式で複数候補を出力する。一般に音素の区分と内容についての曖昧さを表現する音韻(セグメント)ラティス(図-5)の形で出力される¹¹⁾。日本語の場合、音素の種類は約20種であるが、前後の音素環境を考慮して、より多くの種類を区別することもある。

次にハッシュアドレスを用いる一種の連想記憶などを利用して単語候補を限定した後に、Levenshteinの距離などに基づいて文字列間のマッチングを行い、最大類似度になる単語を選ぶ¹²⁾。単語辞書は音素の記号列として表現されている。

(4) ネットワークモデルの利用

単語の音韻構造を変形も含めて効率良く表現するのがネットワークモデルである^{13),14)}。図-6に示すように、各状態には音素記号などが割り当てられる。状態間の遷移を表すアークのつけかたにより、ラベル間の遷移を制御し、発声の変形(無声化など)に対処す

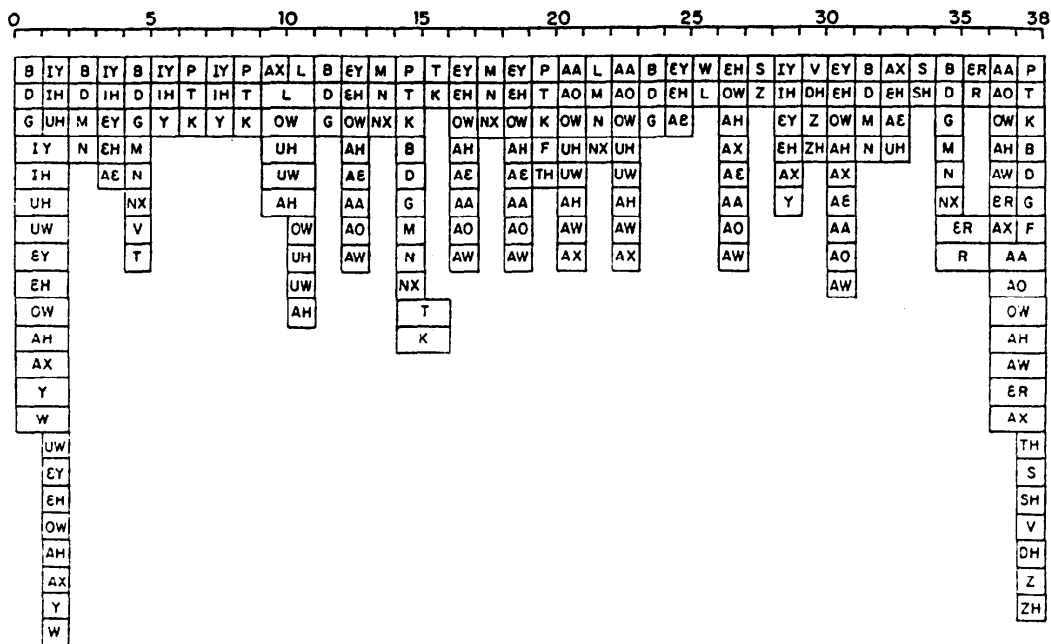


図-5 セグメントラティスの例
発話: "Have any people done chemical analyses on this rock?"



図-6 単語のネットワーク表現例 単語「イチ」の例

る。アークには遷移確率が付与される。認識処理では、入力データの各フレームのデータと各状態のパターン間の類似度によって入力データが観測される確率を算出し、遷移確率も含めて、入力スペクトル系列に対し最も大きい確率の値を計算し、これをその単語とのスコアとする。最終判定は、一番スコアの大きい単語を選ぶことにより行われる。

またネットワークモデルは単語内の構造のみならず、文の構造も単語の並びの関係をネットワーク表現できるので、連続音声認識における入力文のモデルとして使われている。連続音声の場合、語彙数 1011 の比較的限定した文しか許さないタスク（情報検索）でも状態数は 15,000 と報告されている¹³⁾。このようなネットワークはリスト形式のデータ構造によって表現できる。

(5) 音声理解

多くの知識源を利用して発話の内容を理解するシステムは音声理解と呼ばれる。音響・音素・単語・構文

情報・意味情報といった多くの知識源が利用されることが特徴である。この中で、音響分析・音素識別は単語認識の場合と同じである。また単語レベルでは、単語の位置が不明確になってくるので、音韻ラティスと同様に単語ラティスの形式で表現される。

言語情報として構文情報と意味情報がある。構文情報は文の単語構成を表すものであり、文脈自由文法や格文法が利用される。また単語の接続を 2 単語のペアについて単語接続テーブルの形式で用いることがある。

意味情報は、意味素性（セマンティックマーカ）により単語の基本的な意味を表現したり、セマンティックネットワークにより単語間の関係を表現する。

音声理解の場合は、知識源を利用するシステム構成が大きな問題となる。図-7 に黒板モデルと呼ばれるシステム構成を示す。これは音声理解研究で Hearsay II¹⁵⁾ というシステムで開発された考え方であり、黒板は言語のレベルに応じて場所を区分しておく。知識源である構文・意味などの処理を行う知識源の各モジュールは、黒板を観測することにより起動され、その出力結果を仮説として、黒板上の定められた位置に書き込む。このような非同期・並列的な動作により、入

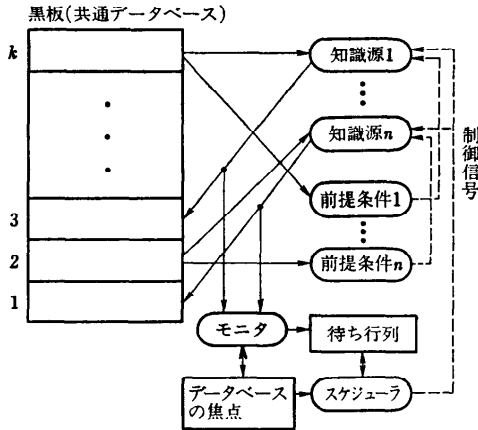


図-7 Hearsay II の黒板モデル

力された文音声の内容を判定する。この場合、システム全体の処理速度を決定するのは、作業記憶である黒板を構成するメモリの能力となり、マルチポートアクセスとともに高速なものが望まれる。

3.2 音声合成

音声合成システムは用途により異なるが、できるだけ小さな構成で自然性の高い明瞭な音声合成できる方式を利用することが望ましい。それぞれの方式の特徴を生かした種々の用途が考えられている。

システムをコンパクトにまとめるには、方式の特徴に応じたデータ構造を採用する必要がある。

(1) 録音編集方式

基本的には波形そのものを記憶する方式であり、データ圧縮率の高いいろいろな符号化方式が考えられている。この方式の特徴は、自然音声を記録するので明瞭性・自然性が秀れている反面、記憶容量が大きいという欠点がある。そこで、出力するメッセージ間で共通で、できるだけ大きな部分を単位として記憶できるデータ構造をとる。たとえば、駅構内の案内放送では、

- 1 番線に、上り列車が参ります。
- 2 番線に、上り列車が参ります。

などのメッセージがある。そこで、図-8 に示すようなデータ構造で音声を記憶する。ここでの記憶単位は、「1 番線に」、「上り列車が」、「参ります」などのメッセージ間で共通に利用できる文節である。メッセージ文として合成するには、それぞれの文節音声を繋ぎ合わせなければならない。このため、図-8(b)のような、メッセージ文と接続する文節の順序との関係記憶するテーブルを用意する。出力に際してはメッ

アドレス	文節 No.	内 容
A ₁	文節 1	いちばんせんに
A ₂	文節 2	にばんせんに
A ₃	文節 3	のぼりれっしゃが
A ₄	文節 4	くだりれっしゃが
A ₅	文節 5	まいります

(a) 文節データ

セグメント	文番号 (1) (2) (3)			文節 No.	メモリアドレス
	(1)	(2)	(3)		
(1)	1	1	2	1	A ₁
(2)	6	6	6	2	A ₂
(3)	3	4	4	3	A ₃
(4)	5	5	5	4	A ₄
(5)	7	7	7	5	A ₅
6:	ポーズ			6	A ₆
7:	終了				

(b) データ構造

図-8 録音編集方式のデータ構造の例

セージの文番号を指定すれば良い¹⁶⁾。

メモリに記憶する音声波形データは、PCM、ADPCM、APC などの高効率の符号化方式でデータ圧縮が行われる。したがって、復号器により音声波形に変換する必要がある。この方式のメモリ容量は、1 秒当り 12K ビットから 64K ビットである。

駅の案内放送のようにメッセージ文が限定されている用途では、音声データの記憶用メモリとして ROM が適している。留守番電話のように、時によりメッセージが変わる場合は、RAM や磁気ディスクが必要となる。

(2) 分析合成方式

分析合成方式は、録音編集方式の音声波形の代わりに、合成フィルタを駆動するパラメータを記憶するものである。分析方法には、PARCOR, LSP, CEPSTRUM などがある¹⁷⁾⁻¹⁹⁾。たとえば、PARCOR 方式では PARCOR パラメータ K_n による漸化式 (この演算を行うのが合成フィルタ) で合成処理が行われる。この方式では K_n パラメータの他、有声音/無声音のスイッチ、有声音の場合に声の高さや強さを与える声帯音源のパルスの発信周波数 (ピッチ周波数) とそのパルスの振幅、無声音の場合の音源である白色雑音の振幅などを与える必要がある。

表-1 に TI 社の Speak and Spell の合成フィルタの各パラメータのビット数の割り当てを示す^{20), 21)}。ここでは容量の削減を行っている。合成フィルタへのパラメータの供給は 20ms ごとに行う。したがって、

表-1 PARCOR 方式の各パラメータのビット数の割り当て

パラメータ	ビット数
音源振幅	4
ピッチ	5
K_1	5
K_2	5
K_3	4
K_4	4
K_5	4
K_6	4
K_7	4
K_8	3
K_9	3
K_{10}	3

1秒の音声合成には2,400ビット必要となる。しかし、定常的な部分ではパラメータの値はあまり変化しないので、繰り返しビットを用意してこのビットの値により前のパラメータを引き続き利用できるようにしている。この結果、平均的に約1,200ビット/秒のビットレートとなっている。

分析合成方式は、音声の生成機構をモデル化しており、合成フィルタのパラメータのほかに、音源のパルス波形も音質に影響する。PARCOR方式は原音声をモデルで近似するので、分析時に近似誤差が生じている。この誤差は、声帯の振動による音源パルスの時点が大きくなる。そこでこの誤差波形をROMなどに記憶して、合成フィルタの音源として用いる場合もある²²⁾。

この方式では、良質のパラメータを短時間で求めるのは簡単ではない。そのため、あらかじめ分析されたパラメータを利用する場合がほとんどである。したがってROM型のメモリをよく用いる。記憶容量は1秒当りの音声に1.2Kビットから9.6Kビット要する。

(3) 規則合成方式

録音編集方式や分析合成方式は、基本的には自然音声を記憶して再生するものである。一方、規則合成方式は、自然音声のごく一部を音声素片として利用する。素片として記憶されるものは、子音(C)と母音(V)の組み合わせであるCV音節やCVC、VCV音節である。このような小さな単位を接続することで任意の文の合成が可能となる。素片は文字で言えば仮名に相当するもので、メッセージとして構成するにはこれらを接続しなければならない。また、聞きやすい音声として接続するためには、個々の音の長さや抑揚、単

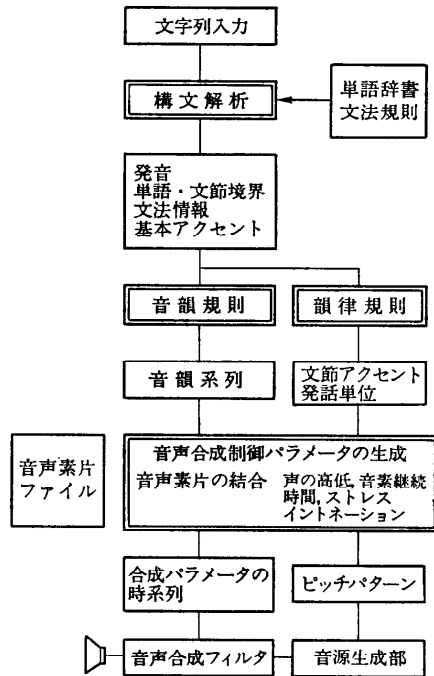


図-9 規則合成方式の処理の流れ²³⁾

語としてのアクセント、文としてのイントネーションなどがよく制御されている必要がある。規則合成方式では、音の接続の方法を規則として用意している。こうすることで、任意のメッセージ音声の合成が可能となる。

図-9に一般的な規則合成方式の処理の流れと、各処理で利用される規則を示す²³⁾。規則合成方式の音声合成装置は、出力する文を文字列として入力する。そのため、文を構成する単語を見出すための単語辞書、単語の接続の仕方を規定する文法情報、各単語の発音・アクセントなどを規定する音韻規則、文のイントネーションや文中のポーズの位置、長さなどを与える韻律規則などを用意しなければならない。たとえば、「この音声は、規則合成方式で合成されました。」のようなメッセージを音声合成するとした場合、この文は、

この、音声、は、規則、合成、方式、で、
合成され、まし、た。

などの単語から構成されている。入力文字列から個々の単語を検出し、その発音の仕方を与え、アクセント、文としてのイントネーションなどを付与して聞きやすい音声メッセージとするのが規則の役割である。

最終的な音声の合成は、対応するCVやCVCな

どの音声素片のパラメータを接続して、合成フィルタを駆動することで行われる。音声素片数は、CV 音節で 100 から 200 程度、VCV 音節で 700 から 800、CVC では数千となる。

4. 音声認識・合成システムの今後とメモリへの期待

音声は、音声生成器官の運動により生成される。器官の各部は質量を持つため、運動は連続した滑らかなものとなる。したがって、原理的に個々の音を正確に区切ることは困難である。音声認識においては、この問題は調音結合として個々の音の認識を難しくしている。また、音声合成では、滑らかで自然な音声をいかに生成するかの問題を投げ掛けている。

音声認識技術として、この問題の解決策の一つは言語情報の利用である。すなわち、言語としての知識を利用して、あり得ない音を除いたり不足している音を補ったりするのである。この処理を実時間で実行するには、きわめて高速のプロセッサが必要であり、したがってメモリも高速アクセスの可能なものが要求される。なぜなら、ある時点で可能な組み合わせは数多くありそのすべてを調べる必要があるからである。また、ある文脈で使用される単語や構文は制約を受けている。そこで、その時点までに明らかになった情報により、単語の候補や構文を限定することも有効であろう。ここで、メモリに期待されるのは、このような一種の連想処理が容易に実行できる機能を持つことである。従来、計算機のメモリ内で論理的に組み立てられていたデータ構造を物理的なメモリの構造としても、持つようにできれば便利であろう。このようなメモリは、音声認識や合成処理だけでなく人工知能の処理にも有効である²⁴⁾。

音声合成技術では、調音結合の問題は音声生成のモデルの改良が一つの解決策である。このためには、膨大なパラメータにより合成装置を制御しなければならない。そこで現実的には、それぞれの調音の仕方に応じた音声素片を数多く用意し、場合により使い分ける方法が考えられる。この場合のメモリとしては、同一の記号でも前後の記号列により検索されるものが異なる機能が必要になる。

音声認識の項で解説したように、知識源を利用した高度な処理が今後要求されるようになる。このとき、一つの内容を複数のプロセッサがアクセスする必要がある。現在のメモリシステムでは、このような場

合、衝突を避けるためいろいろな工夫がされている。構造的に衝突の心配のないメモリも期待される。

音声認識システムは基本的に入力装置であり、対応するものとして、スイッチ、キーボードなどの安価なものに対抗しなければならない。この意味で、LSI 化は避けられない方向である。このとき、処理機能とメモリとが一体化される Logic in Memory の出現も待たれる所である。

本稿では音声情報処理技術における各種のデータ構造とメモリについて整理した。音声情報処理技術は現在発展中のものであり、確立された技術は一部である。そのため、十分に解説できたとは言いがたいが、今後のメモリ研究に参考になれば幸いである。

参 考 文 献

- 1) 新見康永：音声認識，共立出版（1979）。
- 2) 藤崎博也：最近の音声合成技術の動向と展望，電子技術，Vol. 22, No. 1, pp. 9-13（1980）。
- 3) 伊勢孝雄：DECtalk, ASCII, Vol. 9, No. 3（1985）。
- 4) 板倉文忠，東倉洋一：音声の特徴抽出と情報圧縮，情報処理，Vol. 19, No. 7, pp. 644-655（July 1978）。
- 5) Vintsyuk, T. K.: Speech Recognition by Dynamic Programming, Kibernetica, Vol. 4, No. 1, pp. 81-88（1968）。
- 6) 迫江博昭，千葉成美：動的計画法を利用した音声の時間正規化に基づく連続単語認識，日本音響学会誌，Vol. 27, No. 9, pp. 483-490（1971）。
- 7) Vintsyuk, T. K.: Element-wise Recognition of Continuous Speech Composed of Words from a Specified Dictionary, Kibernetica, No. 2, pp. 133-143（1971）。
- 8) Sakoe, H.: Two-Level DP-Matching - A Dynamic Programming-Based Pattern Matching Algorithm for Connected Word Recognition, IEEE Trans. Acoust. Speech, Signal Proc., Vol. ASSP-27, No. 6, pp. 588-595（1979）。
- 9) 千葉，亘理，渡辺：音声認識システム，大型プロジェクト「パターン情報システム」研究開発成果発表会論文集，pp. 157-165（1980）。
- 10) 麻田，坂田，高橋，竹林，平井，篠田，新田，渡辺：ハイブリッド構造マッチング法による電話音声の認識，音講論 1-4-16（Mar. 1982）。
- 11) Woods, W. A.: Motivation and Overview of SPEECHLIS - An Experimental Prototype for Speech Understanding Research, IEEE Trans. Acoust. Speech, Signal Proc., Vol. ASSP-23, No. 1, pp. 1-10（1975）。
- 12) Kohonen, T., Riittinen, H., Jalanko, M., Reuhkala, E. and Haltsonen, S.: A Thousand-

- word Recognition System Based on the Learning Subspace Method and Redundant Hash Addressing, Proc. 5-th IJCP, pp. 158-165 (1980).
- 13) Lowerre, B. and Reddy, D.R.: The Harpy Speech Understanding System, in (31) pp. 340-360.
- 14) Jelinek, F.: Continuous Speech Recognition by Statistical Methods, Proc. IEEE, Vol. 64, No. 4, pp. 532-556 (1976).
- 15) Erman, L.D., Hayes-Roth, F., Lesser, V. R. and Reddy, D.R.: The Hearsay II Speech Understanding System: Integrating Knowledge to Resolve Uncertainty, ACM Computing Surveys, Vol. 12, No. 2, pp. 213-254 (1980).
- 16) 奥田直紀: ADPCM 信号に変換して ROM に書き込む電子化音声応答装置, 電子技術, Vol. 21, No. 12, pp. 56-59 (1979).
- 17) 板倉文忠: スペクトル符号化に基づく音声分析合成, 日本音響学会誌, Vol. 37, No. 5, pp. 197-203 (1981).
- 18) 板倉文忠, 嵯峨山茂樹: 線スペクトル周波数をパラメータとした音声合成法とその LSI 化, 日経エレクトロニクス, pp. 128-158 (Feb. 1982).
- 19) 斉藤取三, 中田和男: 音声情報処理の基礎, オーム社 (1981).
- 20) ウィギンス, R., ブランティンガム, L.: 線形予測符号化 (LPC) ボコーダの音声合成装置を 3 チップで実現, 日経エレクトロニクス, pp. 147-162 (Jan. 1979).
- 21) 3 チップの LSI で作られた音声合成学習機 Speak and Spell, 電子技術, Vol. 21, No. 12, pp. 63-64 (1979).
- 22) 楠木好明: 音声合成 LSI, MN 6401/MN 1261, オートメーション, Vol. 26, No. 10, pp. 42-47 (1979).
- 23) 古井貞照: デジタル音声処理, 東海大学出版会 (1985).
- 24) Stefik, M.: Strategic Computing of DARPA: Overview and Assessment, Comm. ACM. Vol. 28, No. 7, pp. 690-704 (1985).
- 音声情報処理の一般的な参考書・特集は, 1), 19), 23) のほかに次のようなものがある.
- 25) 電子通信学会編: 聴覚と音声, 電子通信学会 (1966).
- 26) 大泉充郎監修・藤村 靖編著: 音声科学, 東京大学出版会 (1972).
- 27) 中田和男: 音声, 日本音響学会編, 音響工学講座, コロナ社 (1977).
- 28) 特集 Man-Machine Communication by Voice, Proc. IEEE, Vol. 64, No. 4 (1976).
- 29) 特集 Man-Machine Speech Communication, Proc. IEEE, Vol. 73, No. 11 (1985).
- 30) 特集 音声情報処理, 情報処理, Vol. 24, No. 8 (Aug. 1983).
- 31) W. A. Lea (Ed): Trends in Speech Recognition, Prentice-Hall (1980).
- 32) 中川聖一: 音声情報の認識と理解, 坂井利之編著, 情報基礎学詳説, コロナ社, 第4章 (1983).
- 33) 小林 勉: 音声合成・認識用 LSI, 日本音響学会誌, Vol. 39, No. 11, pp. 744-749 (1983).
- 34) 任意の文章を音声に変換する規則音声合成が実際に, 日経エレクトロニクス, pp. 121-142 (July 1984).

(昭和 61 年 4 月 9 日受付)