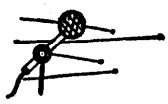
講　演



情報システムにおける音声の認識と合成†

坂　井　利　之††

1. はじめに

これから約1時間『情報システムにおける音声の認識と合成』という標題でスライド40枚，録音テープ4分30秒を用いてお話し致します．最初与えられたテーマは『音声の認識と合成』ということでした．しかしこのテーマではすでに雑誌や講演などでも数多くの報告がありますし，また，情報処理学会は音声学会でもなく音響学会でもありませんので，音声を何が何でも主題とする立場ではなく，符号，文字，画像など人間やコンピュータで用いている他の情報媒体と対等の位置づけで扱って行きたいと思いまして，標題のように致しました．したがって今日の話は，勿論忠実なサーベイでないばかりでなく，危険を承知の上で独断的な判定も予測も申し上げます．お話の順序はスライドに示しました通りでありますが，私は今日の話を応用面と利用空間について，それぞれ3つずつに分類することで形を整えたいと思っています．まず音声認識・音声合成の応用面については 表-1 に示しますように(Ⅰ)電気通信用，(Ⅱ)マン・マシン・コミュニケーション用，(Ⅲ)データベース用と3大別致しました．現在実用に入りだしてきたのは認識でも合成でも(Ⅱ)の初期であります．(Ⅲ)は研究，開発の段階にあるといえましょう．次の3分類は 表-2 のように音声情報処理を“コ”(話者)の空間，“ソ”(話者・聞き手)の空間，“ア”(話者・聞き手・第三者&環境)の空間に区別することです．“ア”の空間における音声認識の必要性と技術をお話することが，私の今日の話のポイントです．これが展開致しますと，先程の Evans さんの話に対応して言うならば，音声の非電話系の処理技術が現在の数十億円程度のマーケットから，数千，数百億の産業に発達するのではないかと思っています．

† 情報処理学会第21回全国大会招待講演（昭和55年5月22日）
†† 本会前副会長　京都大学工学部情報工学科
* これは(Ⅱ)から(Ⅲ)への橋渡しで将来はマン・マシン・コミュニケーションの主流(Ⅲ)はとなります．

2. 音声認識，合成の位置づけ

(a)　音声の応用面

表-1に音声認識，合成の応用面を示します．(Ⅰ)の電気通信用というのは音声を通信系の立場からとらえ，特徴パラメータのレベルで分析・合成を行うものであり，分析・合成が1つのペアを形成しています．音声情報量の圧縮が可能であり，人間に対する音声情報としての品質は保存されていますが，人間にとっても機械にとっても音声が何であるか判っていないのが特徴であります．

(Ⅱ)のマン・マシン・コミュニケーション用というのは，表にも示してありますように，人間と機械が対話する立場での音声の応用面であります．現在のところ，座席予約とか，銀行の残高案内のように，会話の内容が定型的で単純で，伝票の項目を埋める形のものが実用になっております．この場合の音声認識は単純なタスクの音声理解ということができます．これらに対し会場案内システムでは，今の例のように，定型的で単純な会話文章でなくて，データベース（知識）の利用*が必要となります[1]．たとえばデパートで売場を尋ねる時のように，質問者のいる階(売場)，買いたいもののある階（売場）の組合せはきわめて数多く，しかもデパートでは売場の変更が常ですから，知識（データベース）にコンサルトする必要があります．また質問音声の認識は代表的な質問文の理解で，実体は文章中のキーワード認識となります．合成は，質問者のいる売り場から見える目標物を示しながら，階段，エレベータ，エスカレータなどに到達するよう誘導し，目標の階と売り場に行けるよう案内する文章を合成することになります．座席予約などと内容的に異なることが判って戴けると思います．

(Ⅲ)のデータベース用というのは，音声だけの用途を考えず，文字との併用が最初から意図されたシステムであります．マン・マシン・コミュニケーションの情報媒体として文字，音声の選択が人間の指示により

表-1 音声認識・音声合成の応用面

(I) 電気通信用	電話 データ圧縮通信 (パラメータ伝送)	音声 & 音声 (人間) (人間)
(II) マン・マシン・コミュニケーション用	質問・応答系 (座席予約システム) (銀行残高案内システム) (会場案内システム)	音声→音声理解〔単純タスク〕 (人間) (機械) 音声編集合成〔単純応答文〕→音声 (機械) (人間)
(III) データベース用 (情報検索)	文字・音声併用の 情報ネットワーク (音声入力システム) (音声応答システム)	音声→音声認識・理解 (人間) (機械) 音声法則合成→音声 (機械) (人間)

を位置づけることが大切であると思います．情報システムという標題からは，大きいシステムを想像されるかも知れませんが，これは情報媒体を対等に考えるという意味でありまして，stand alone の家電，農業機械を考えても少しも構わないわけです．単体機器では，音声のもつ流行性，方言性が活用できます．今月の機械の声は○○さんの声にしようと決めることもできますし，代理店や現場で方言の音声合成を機械に入れることも可能となれば音声媒体の有利さが発揮できます．

場面に応じて切換可能であります．音声理解，音声合成いずれの面でも最も一般的・汎用的なものといえましょう．

音声の一番よい点は手にマイクやスイッチなどを持っていなくても，支配できる空間が数m以上もあることです．この点を活かしますと，家電でも産業用でも新天地が開拓されます．手にマイクを持ってTVなどのチャネル切換えを音声認識でさせていたのでは，本当に音声の持つ特徴を活かしていません．目や手が作業に使われていても，耳から割込みを受付けるのが音声の特徴です．音声の欠点は無責任ということであります．文字の場合には1つの文字ごと，単語ごと，文章ごとに人間にフィードバックができ，構文的にも意味的にも責任のある文章を生成できますが，音声ではそうは参りません．今私の話している一分前は何であったとか言われても，私には判りません．

文字にはハードコピーとソフトコピーがあり，ハードコピーでは複製が可能です．ドット文字になると益益図面の仲間入りをしますが，本来漢字も数式記号，化学記号も図と同じ扱いになります．ソフトコピーはやや低情報密度であり，表示に実時間性が要求されます．図面を扱うファクシミリなどで，文字とグラフを同一視している技術が良くない事が判ってきております．不均質である情報源の構造把握の上に立った処理が大切である[2]ことが，これから述べます音声情報源の構造把握の必要性*と一脈相通ずるものがあります．

濃淡画像は静止画，動画と区別されますが，情報量が大きく，符号1に対し音声 10^3，画像 10^6 であるという膨大さの外に，構造化，シンボル化ができ難いという特徴があります．

コンピュータの出現，超 LSI の長足の進歩で，これら情報媒体間の相互変換が実用的に可能になってきている事実と将来を洞察し，その中で音声の情報媒体

* この文章では省略（文献9）参照）

(b) 音声処理の空間

次に音声通信，音声認識，話者認識，音声合成を利用する空間について考えて見ます．表-2 では“コ”，“ソ”，“ア”という空間の分け方をしました．日本語で主語がよく省略されるのは，日本語が“ア”の空間の言語であるからです．日本語は，送・受の人(話者，聞き手）のほかに，まわりの環境も頭の中に入れて話すのが通常です．“コ”というのは「これ」「この」「ここ」というように話者の空間，“ソ”というのは「それ」「その」「そこ」というように聞き手の空間で，勿論“ソ”の空間では，送・受両方が判っているという立場をとります．通信の例で示しますと，インタビュー，対談などとなります．“ア”というのは「あれ」「あの」「あそこ」というように，送・受両方の空間の外の，第三者の空間も対象の中に入ります．街頭録音で言えば第三者や環境の音も混入してきます．“ア”の空間は Robotics と換言してもよいと思います．Robotics では，送・受二者間において相互干渉がある interactive な communication や対象（受け手）の制御を含むだけでなく，更に環境の制御，送り手の判断や行動をも考察の対象に致します．したがって最も一般的な世界（“ア”の空間）におけるコミュニケーションと行動を扱うのが Robotics であります．

表-2 の音声認識の欄で，“ソ”の空間に書いてある対話継続形音声理解というのは，先程説明致しましたデパート案内で，機械の音声認識の不調が生じたような場合でも，途中で会話が切れないよう配慮したシステムです．one word speech recognition というのは，“危い”“火事だ”“止めて”などという単語を，語調，周囲の環境（騒音，煙，スピード etc.）などを総合的に判断し，平常会話中の同じ単語と区別して手

表-2 空間と音声情報処理

共通に考慮すべき項目 {用途（産業，オフィス，電話，民生）と音声品質（識別スコア）/コスト
ハードとソフトの分離性（変更所要時間，方言の導入，流行性の付与）
音声媒体の特徴の洞察（家電，事務器，情報・制御機器，非常用機器，避難誘導案内，CAI）

空間	通信	音声認識	話者認識	音声合成
"コ"（話者）	講義 放送 （1方向）	限定話者，無響室 接話マイク使用	声紋， 文章，使用単語	録音・素片より合成
"ソ" 話者 聞き手	インタビュー，対談 電話 （双方向）	不特定話者 対話継続形音声理解	用語の対話的指示	質問・応答システム 機器操作指令・警報
"ア" 話者 聞き手 第三者 & 環境	会議電話 街頭騒音 (Robotics) 赤電話コーナー	one word speech recognition speech filter	特別室（カード） 特別端末 併用（サイン，TV）	公衆放送（監視・制御システムより） （非常時：火災，地震，事故）

許に機器もなくSN比の悪い環境で行う音声認識であります．これができると，音声認識は，店舗，ホテル，自動車の中にも応用可能となり，産業規模の様子が変って参ります．このように劣悪な環境での音声認識が成立するためには，人間の声と，声以外の音との区別が必要で，これを行うのが音声情報フィルタです（実例は後で録音でお聞かせ致します）．話者認識の欄で，"ソ"の空間に「用語の対話的指示」と書きましたのは"間"の有効利用であります．

数個の登録単語の中から，その内の1つを話すよう機械から指示されたら，間を間違わずに人間にその単語を話させることによって，本人と，本人の声の録音や偽声とを区別しようとするものです．同じく話者認識で"ア"の空間での応用場面は，IDカードで特別室に入り，特別の人のみ利用可能な端末で本人の資格をオンラインサインまたは話者認識*などで厳重に確認した上で，データベースへのアクセスを許そうというものです．

音声合成の欄で，"ソ"の空間に書いてある質問応答系は，例えば先程説明しました会場案内のような例です．更に同欄の「機器操作指令」の機械からの合成音には簡単に特定の人の声とか流行語，方言などを入れられるシステムも充分考えられるとしています．音声合成・"ア"の空間は，非常時における地下街，航空機，船舶，建物などでの避難誘導案内を考えています．実際に現在生じた事故を知った上での，最も適切な避難誘導を音声で行う方式であります．仮定のシミュレーションではありません．選挙予報でいえば，開票初期の実データを入れての予報とか，防災などでは発火地点，気象条件を入れての消防・警備態勢の問題に該当致します．

初めにも申しましたように，音声情報媒体のもつ性質を生かし，その応用範囲の広さと重要さを考えますと，"ソ"から更に"ア"の空間における音声通信，音声認識，話者認識，音声合成の技術開発が，音声情報処理を巨大な産業分野に育てるものであると思います．

* 話者認識としての技術では，"コ" "ソ" の方が難しい．

3. 音声と画像・文書との認識・合成上の技術的差違

図-1にパターン理解システムのブロック図を示します．"コ"の空間，すなわち機械のみによって，限定された音声パターン（限定単語），限られた文字のみを認識するパターン認識はこの範疇に入ります．整備された環境で，対象となるパターンを，予定の位置で観測し，前処理，特徴抽出を行って特定アルゴリズムによって判定（または拒否）を行うものであります．対象の変形，不特定話者の発話を受容し，枠組を作らずに文字の行の検出，文字の切出しなどの操作を含む柔軟なパターン認識を行うためには，機械の中に対象に対するモデルを構築し，送り手の入力パターンに対する適応制御と受け手の機械の中での標準パターンの正規化，学習を伴いますので，"ソ"の空間になり，これはパターン理解と呼ばれています．パターン認識の次の段階であります．さらには"ア"の空間でのパターン理解は，対象そのもの，あるいは環境の制御，観測機器の移動や制御を本質的に含むもので，Roboticsの世界でのパターン理解であります．これには対象（送り手）に対する知識や，その情報源の構造把握のモデルを受け手の方（パターン理解システム）に持っている必要があるのは"ソ"の空間のパターン理解と同じです．パターン認識や理解の階層性が，シグナル，ラベル・シンボル（パラメータ空間），モデル，タスクに分類できますが，詳しいことは文献[4]を見て下さい．パターン理解のモデルを構築するときに，マン・マシン・コミュニケーションが有効であることは，申すまでもありません．モデルは対象を記述するという科学一般の方法論の根本であると共に，予測を可能にし，選択の順序，優先度を決める上でも有効であって，人工知能の技法[5]のポイントの1つでもあります．

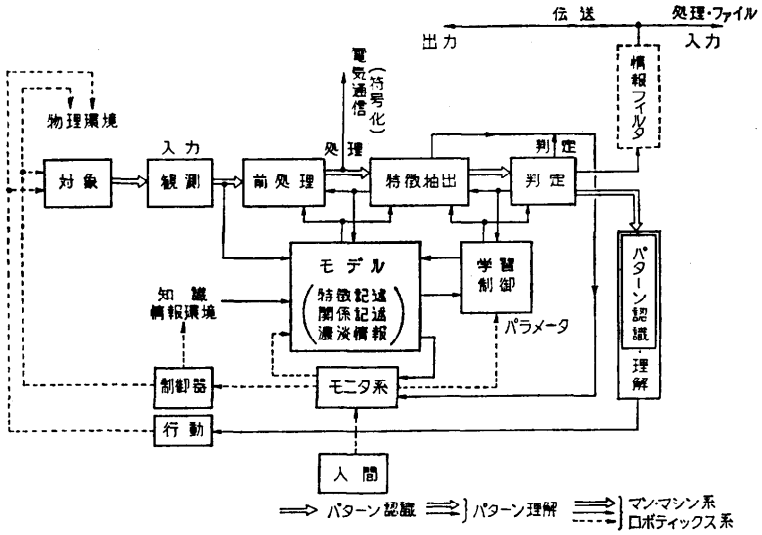


図-1 パターン理解システムのブロック図

図-2 は，音声認識，音声合成の技術的な問題を他の媒体の技術と対比して示そうとしたものです．音声媒体を符号，文字，画像などとの関連で示すと共に，文字読取装置(OCR)，コンピュータネットワークのアーキテクチャにおけるプロトコル[6]の思想などとも対比しました．現在の OCR の問題点は，文字パターンを一挙に符号・シンボルのレベルまで認識しようとする点にあります．これが“コ”の空間，限られたフォントに制限せざるを得ない理由です．音声の場合は，特徴パラメータを豊富に持っていること，このパラメータ空間を，ユーザ空間とパターンという物理空間とのインタフェースにしている点が優れています．文字の世界でも音声のような特徴パラメータ空間を利用することの有効性にふれておきたいと思います．音声応用の3分類（表-1）を図の上で示すと，図-2のようになります．(I)の電気通信用は，音声のパラメータ分析・パラメータ合成から成りたっていて，ピラミッド（階層構造）のパラメータ・レベルまでで閉じたシステムを作ります．(II)のコミュニケーション用は音声認識によって音韻符号または単語まで達する音声認識と，録音音声波形または音声素片の編集合成を行うものであって，図-2において（II）と示したルートをたどります．

(III)のデータベース用は，図にも示してありますように音声の分析を経てデータと示したデータベースに到達する音声理解の部分と，ここに存在する正書法のシンボル系列から逆に音韻列・アクセント記号として示した音韻記号列への翻訳の部分と，音韻列による法則合成とから成りたっています．

合成音声例として，後で録音によってお聞かせするものを図の中で説明致します．図形（波形）制御パラメータからの play back (OVE II)，音韻記号系からの合成，パラメータ空間(PARCOR)での分析・合成，法則による素片編集の合成（零交差波），Text からの法則による合成 (MITALK' 79) であります．その位置づけが判って戴けると思います．

英語と日本語との大きな差違は，音韻列(話し言葉)と正書法（書き言葉）との翻訳ルールの差違でありまして，日本語は，きわめて簡単であります．電話は図にも示してありますように，階層構造のパイプの中を素通りしています．パラメータもシンボルも認識されずに，音声波形のままで送受間を通り抜けています．ユーザ空間の人間と書いた音声言語が，発声器官で音声波形になり，受け手の人間の聴覚器官に波形のままで伝達されています．図のピラミッドのユーザ空間は人間，下の情報媒体，物理パターンは機械により処理されるものです．

4. 音声の認識

音声メッセージの発生のメカニズムを模式的に図示すると**図-3**のようになります．発生において，深層から，より表層に至るまでの順に，心理，意味，談話形

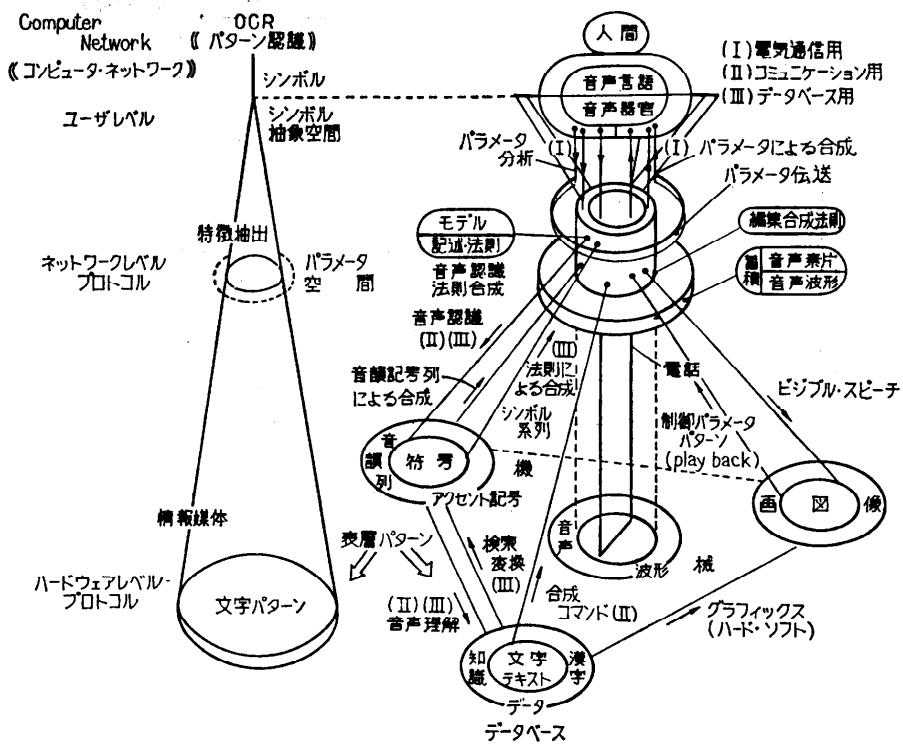


図-2 音声認識・音声合成技術と情報処理技術

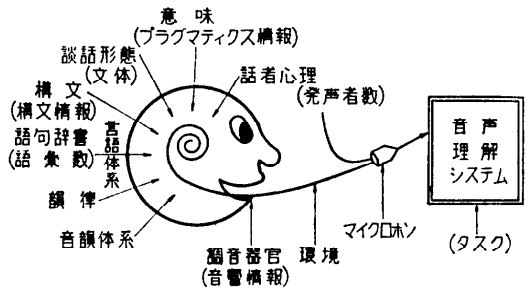


図-3 音声メッセージに内包される多くの因子

態，構文，韻律，調音器官として示してあります．音声理解とは，音声生成の逆の手順をたどってゆくもので，深層の言語や心理に含まれる知識を必要としますので，現在の技術では音声メッセージの範囲を限定し，どのような意図の下に，どのような目的のために発声されたかのタスクを仮定するのが一般的であります．したがって音声理解システム（“ソ”の空間），あるいは音声認識（“コ”の空間）で機械を設計するときの仕様書[7c]に，規定を必要とするものが括弧内に書かれていると考えて戴いて結構です．

表-3にはこれら仕様設定に対しての項目，条件，認識技術のポイントを示してあります．普通の解説文などに認識技術とか方法として説明がしてあるのは，この表に対応するものです[7a]．パラメトリックとしてあるのはモデルを用いる方法ですが，これは上述来のモデルとは異なって，通常特定の統計分布を仮定したものです．ノンパラメトリックな方法は，訓練サンプルパターンを最適に分離するような関数を求める方法であります．線形，非線形，区分的線形のものなどがあります．音声認識装置の一般的な構成は図-4に示すようになっています．これは図-1のパターン理解システムのブロック図を，さらに具体的な音声認識システムの用語で書き直し，処理の技術，処理の階層を記述し，表-3の仕様条件をも記入したものです．これの説明は，音声認識の本からも[8]雑誌解説[SP]からも簡単に得られますので省略致します．

表-3 音声認識・理解の仕様条件

条件	項目	認識技術のポイント	
		処理項目	ポイント
規定有り：単語数 文法 タスク	認識対象の規定	限定語 孤立単語 連続単語 連続音声 文章（自然語） 人工的文章	標準パターンの登録数少し セグメンテーション不要 セグメンテーション要 セグメンテーション要，各階層知識利用 言語知識利用度大
特定—不特定 性別・年齢 方言 協力的／非協力的	話者	話者類別 個人差の正規化 学習の有無 標準（モデル）の記述	標準パターンの登録数 非線形マッチング（DP） 学習サンプル数，教師の有無 標準パターンの記述形態 （単語／音，パターン／シンボル）
部屋の静・騒 マイクの種別 電話系経由	発声環境	SN 比改善法 放射特性補正	マイクの有無・設置場所 スペクトル包絡の変歪
孤立—連続 発声原稿の有無	発声法	セグメンテーション 調音結合 間投詞の検出	誤りの多いシンボル系列 非線形パターン伸縮 韻律情報
音響分析	音響処理	スペクトル分析 線形予測分析 相関分析	分析区間 適応の有無 パラメトリック／ノンパラメトリック 〔モデルの有無〕

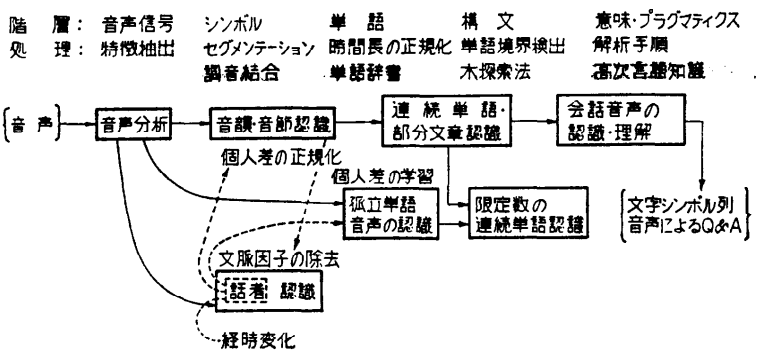


図-4 音声認識・理解システムの階層構造と処理

表-3には，要約して音声特徴パラメータとその特性を記してあります．具体的な定義や特徴は成書[8),BK)]や解説に譲ることに致しますが，文字や図形に比べて特徴パラメータの数が豊富で，それらの抽出技術，パラメータ間の相互関連などが明確になったのは最近10年間で，日本の寄与が多いのも特筆すべきことでしょう．

音声認識において重要だと私が最近思っていることは，音声情報源の構造，すなわち人間の発声器官の構造を十分に考慮した音声認識方式あるいは音声情報フィルタの思想と技術であります．この考慮の有無がかなりキーポイントになるのではないかという実験データを得つつありますが，これについてはすでに一部報告してあります[7a),3)]．今後も数多く報告されると思います．

音声理解システムにおいては，音響レベルでの音声パターン分析から得られるデータ

表-4 特徴パラメータとその特性

特徴パラメータ	抽出の精度	次元（パラメータの数）	認識への寄与の度合い	忠実度(分析・合成系)	処理時間コスト	構築モデルとの対応性	
零交差波	優	5～10	可	劣	優		波形に対し演算操作
Walsh Hadamard 変換	優	>20	可	可	優		
スペクトル (FFT)	優	>20	可	優	良		
自己相関関数	優	5～20	良	優	良		
線形予測係数	優	8～12	劣	優	良		
偏自己相関係数 (PARCOR)	優	8～12	優	優	良		
ケプストラム	優	8～12	優	優	可		
滑らかなスペクトル	優	>20	優	優	劣	良	発声機構のモデル
フォルマント	良	3	優	良	劣	優	
声道断面積関数	可	8～11	優	優	良・劣	優	
唇・舌・アゴの位置	劣	3～6	良	可	劣	優	

表-5　種々のタスクにおける利用した知識と認識率の関係

	タスク		数字	算術文	カレンダー	計算機網
語彙数			10	24	30	101
発声者数			20	5	12	10
入力文数			—	80	120	200
入力文中の単語数			1,500	560	647	1,983
利用した知識	音響情報	文認識率(%)	—	20	8	—
		単語認識率(%)	97	78	54	—
	音響・構文情報	文認識率(%)	—	71	40	58
		単語認識率(%)	—	94	83	91
	音響・構文・プラグマティクス情報	文認識率(%)	—	—	70	64
		単語認識率(%)	—	—	90	93

タスク：数字〔数字のみの発声，データ単独の入力；構文，意味情報なし〕
算術文〔数字（1万まで）の四則演算：例 (12+345)÷6.7；( ) のペア，"+−×÷"の前後は数字，"・"の前後は数字〕
カレンダー〔月・日・曜日：月日と曜日とに一対一対応あり；月は1より12まで，日は1より31まで〕
計算機網〔日本語単文という構造のみ，単語数は101ケ，平叙文・疑問文・命令文の形あり，10話者（通常発声速度と態度）〕

の外に，構文の知識，プラグマティックスや意味の情報が利用できます．知識の利用は，その知識がなければ不明確で決定できないことが可能となり，きわめて多く存在する候補を著しく少なく制約してゆく効果があり，判定結果，信頼性の向上でも，処理時間の短縮でも大いに役立ちます．タスクを設定すること，使用する単語を限定すること，話者を特定の人にすることなどは，それぞれ音声メッセージの可能な構文数を限定すること，未知音声パターンから判定する単語を予測したり強制的に判定することに役立ちますし，また話者を知ることは標準パターンの認識機械へのストアと，未知のパターンとの類似度判定に極めて有効に作用します．

表-5に音響処理部の能力を同一にしたままで，タスクを種々変えた場合の認識データを示します．数字認識のタスクでは，音響処理だけが寄与し，算術文，カレンダーのタスクでは，表の注に示したような構文知識が利用できます．計算機網は最も複雑なタスクであって，構文のみならず，プラグマティックス情報も利用しています．知識によってどのように予測が成立し，選択すべき候補がしぼられるかの具体的な例題は省略致しますが，文献[5]には詳しく説明してあります．

5. 音声の合成

(a)　音声の合成方式

音声の合成の歴史は古く，最初は音響的に，革製の筒などで共鳴系を作って可変にしていました．その後，電気回路で無声音に対しては白雑音源，有声音に対してはパルス発生源を用意しまして音源とし，帯域フィルタでスペクトル分析をして音声を表現することが考えられました．各バンドの振幅は鍵楽器のように指で鍵盤を制御することにより，またピッチはペダルで調製する方式の Voder と呼ぶものが合成の初期のものであります．共に人間の発声器官の構造や機能を模倣したものでありまして，Voder は特徴パラメータとしてスペクトルを採用した分析・合成系の合成器と見ることもできます．事実，人間が適当に演奏していた制御情報を，実際の音声の音響分析から自動的に抽出する方法が，同じく Dudley によって考案されまして，Vocoder と名付けられました．分析・合成系として音声の狭帯域通信研究の嚆矢となったものであります．この方式は表-4に示しました他の特徴パラメータで種々実現されています[7b)e)]．

音声生成モデルによる合成を図-5に示します．音源は上と同じですが，声道の音響伝達特性の生成モデルで2つに分かれます．その1つは電気回路の共振，反共振回路で等価な周波数スペクトルが得られるように作成するターミナル・アナログモデルであります．図-5において，〔 〕内の時間的タイミング，数値としてのピッチ周期，スペクトルパラメータ，強度などのスペクトル合成の制御情報を分析系から得て合成するときは分析・合成系となります．音韻記号列から法則により合成されるときは法則合成[7d)]となります．このモデルには，スペクトルパラメータ，強度などを連

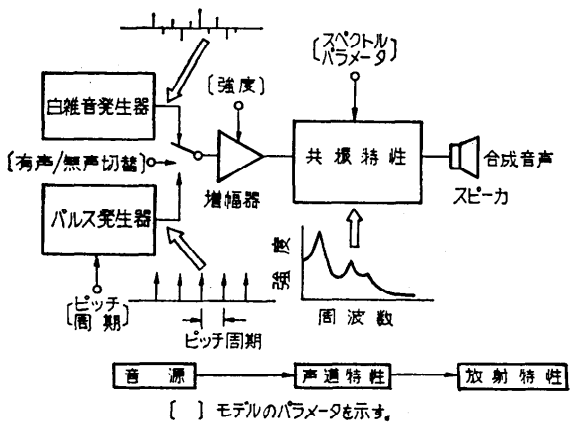


図-5　音声の生成モデル

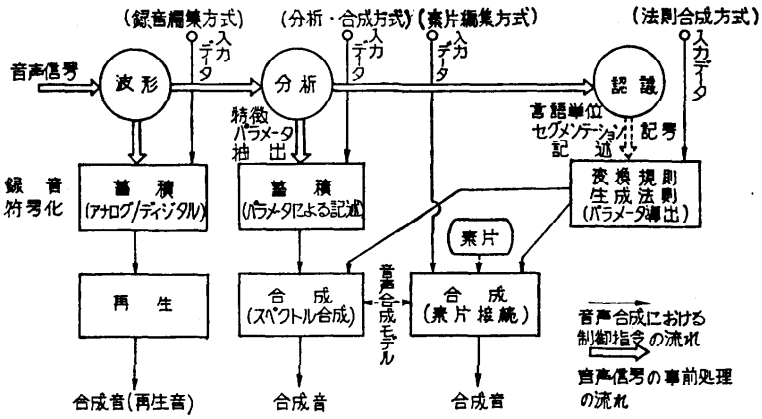


図-6　音声合成方式の分類と階層性　入力データは単語または文章型式

続的な包絡パターンで表示するのでなく，ホルマント型とよばれるモデルをもち，共振周波数と帯域幅が可変の電気回路を数個用いて，音声のスペクトル包絡特性を記述する方法があります．もう1つの生成モデルは，声道も10数個の円形断面をもつ音響管の縦続接続で近似する声道アナログモデルであります．また別の考え方として，人間の音声が共鳴に基づく母音を中心としているものならば，それの一周期分のピッチ周期の波形を記憶させておいて，これを繰返して母音部分を合成し，過渡部の子音部もそれぞれ記憶させた波形の編集によって合成しようとするものがありまして，これを録音編集あるいは音声素片編集の方式といいます．音声合成方式は表-6のようにまとめられます．

音声の合成方式の評価に対しては，文字生成と比較すると類似点が多く，よく判ります．すなわち大きくは音声合成のためのエレメントと，エレメントの合成法則のいずれに負荷をかけるか，換言しますとハードあるいは空間としてのメモリと，ソフト（法則）あるいは制御手順として時間への，システム上からの配分の問題になります．文字の場合について言いますと，手書き文字，あるいはCRTでの文字生成の素子はただ1個で，点であります．この，点の動かし方のルールはしたがって最も複雑であり，スピードは遅く，文字生成の一極端の例であるといえましょう．他の極端は，活字フォントすべてを揃えただけのエレメントを用意し，それらを一列に並べるための選択操作と，その並べる位置の指示操作がルールのすべてである場合です．したがってこの方法はエレメント（メモリ数）が多く，ルールが最も単純でありまして，時間的には

表-6　a）音声合成方式の分類

方式		合成単位	記憶形式
編集型	録音編集型	単語，文節	音声波形
	素片編集型	音節，単音 1ピッチ波形	音声波形，零交差波，インパルス応答
合成型	分析合成型	単語，文節	周波数スペクトル包絡情報 帯域声波器 ホルマント 線形予測係数 PARCOR係数 声道断面積 駆動音源情報
	法則合成型	音節，単音	

b）音声合成方式の比較

合成方式	語彙数	品質	回路	記憶容量
録音編集型	×	○	○	×
素片編集型	△	△	○	△
分析合成型	△	○	△	△
法則合成型	○	×	△	○

高速になります．

音声の場合は，調音結合があり，文字のように時間的，空間的にディジタルに分離することができませんので，合成問題は文字生成より遙かに複雑であります．したがって音声の合成の法則を求めること自体が，音声全般の研究とかなりの範囲で一致するほど複雑でありまして，図-2で言いますと，シンボル系列による法則合成，モデルの記述法則，音声合成法則は，汎用音声認識(Ⅲ)と表裏の関係にあることがわかります．

図-6には音声の階層性（図-2参照）によって合成方式を分類したものを示します．

（b）代表的合成音のデモ

これからお聞かせ致します合成音は，方式の代表と

いうわけではありません．年代，方式，研究所を考慮して，手許にあるものだけでなく，特にお願いして載いたものもあります．

図-2 を参考にして合成の位置づけを考慮した上でお聞き下さい．

（i） Bell Telephone Lab. (USA), 1960年 (L. J. Gerstman) Synthesized Speech on 7090 from Discrete Phoneme Input

図-2 符号より法則(Ⅲ)のルート；音韻列を入力；法則はきわめて単純；少数例；『文章及び歌』

（ii） Royal Institute of Technology (Sweden) 1960年前後 (G. Fant) Synthesized Speech by Öve II, 〔ICA II, NAS Congress〕

図-2 "図"より制御パラメータのルート；Cut and Try で制御パラメータ曲線を入力；手造りで"音韻"の合成側からの研究；少数例；『各国語の母音，文章』

（iii） 京都大学 1968年（坂井・大谷）音声素片(零交差波)を用いた法則による素片編集合成方式[9]

図-2 シンボル系列（ローマ字）より法則による合成(Ⅲ)のルート；任意の音声合成（アクセントなし）可能；英文和訳の音声出力として利用；多数例；『コンピュータによる合声音声の説明，紅葉ゆる……の歌』

（iv） 電電通研 1978年（斉藤） 日本音声言語医学会総会特別講演[10]

図-2 パラメータ分析・パラメータによる合成のルート；『PARCOR 形分析・合成音；原音声：64 kb/秒相当音及び 4.8 kb/秒相当音』

図-2 シンボル系列（ローマ字およびアクセント記号，休止記号，結合記号）より法則による合成(Ⅲ)のルート；VCV 音節単位による法則合成『短文章』

（v） MIT (USA) 1979年 (J. Allen Group) MITALK '79 (Speech Synthesis from Text)[11]

図-2 文字・テキスト・(データベース)より符号(音韻列・アクセント記号）さらに法則による合成(Ⅲ)のルート；任意のテキストより音声の法則合成（ターミナル・アナログ）出力可能；『Synthesized Sentences』

（vi） 京都大学 1980年 〔雑音除去は電電通研，電波研，東大などで実験中〕

音声情報フィルタ；少数例；『短文章の原音声とその音声情報フィルタ通過後音声；"単語音声十パルス雑音"の入力前と通過後；話者より 3 m の距離にてマイク内蔵カセットで録音した短文章の音声；音声情報フィルタ入力前と通過後』

参 考 文 献

1) 石川，浮田，中川，坂井：音声による質問・応答システムの構成―会場案内システムの場合，情報処理学会第 21 回全国大会 2H-10 (昭和 55 年 5 月)．

2) 村尾，坂井：文書画像における構造情報の抽出，情報処理学会第 21 回全国大会，7H-1 (昭和 55 年 5 月)．

3) 坂井，中川：単語音声汎用認識装置の開発，文部省科研費成果報告，Ⅲ 第 4 章(昭和 55 年 3 月)．

4) 坂井：1980 年代の画像処理，情報処理 Vol. **21**, No. 6, pp. 639—644 (1980)．

5) 坂井：人工知能―特に音声画像理解を中心として，bit No. 6, pp. 75—82 (1979)．

6) 電子通信学会：電子通信ハンドブック，第24編，p. 1458.

7) 電子通信学会関西支部専門講習会：音声・画像情報の処理技術とその応用．昭和 55 年 1 月
   a) 坂井：音声認識の基礎，p. 4, p. 6, p. 8, p. 11, p. 12, p. 16.
   b) 斉藤収三：音声合成の基礎，p. 24, p. 28.
   c) 千葉成美：音声認識の応用，p. 32.
   d) 藤崎博也：規則による音声合成，p. 39, p. 40.
   e) 小池恒彦：音声合成（分析合成）の応用，p. 45, p. 50, p. 52.

8) 新美康永：音声認識，共立出版(昭和 54 年 10 月)

9) 坂井，大谷謙治：零交差波による音声合成システムについて，電子通信学会論文誌，**52**-C (1969, 11)．

10) 斉藤収三：音声の合成と認識(音声言語医学，**20**, No. 2, pp. 41—49 (1979, 2)．

11) J. Allen : Synthesis of speech from unrestricted text, Proc. IEEE **64**, pp. 433-442 (1976)．

SP：雑誌特集号

電子通信学会創立 50 周年記念全国大会：シンポジウム S. 11，音声合成，昭和 42. 10.

電子通信学会誌：音声特集号，**51**, 11 号，昭 43. 11.

日本音響学会誌：音声特集号，**34**, 3, 1978. 3.

数理科学：音，No. 116，昭和 48. 2.

日本音響学会誌：音声特集号，**31**, 3, 昭 50. 3.

情報処理：音声と情報処理小特集，**19**, 7, 1978.

エレクトロニクス：音声応答技術特集，音声認識・合成理解の基礎用語，1979 年 9 月.

電子技術：実用化音声認識・合成装置特集，**21**, 12, 1979 年 11 月.

計測と制御：音と計測制御小特集，**19**, 3, 昭 55. 3.

Proc. of IEEE **64**, No. 4, Special issue on man machine communication by voice, 1976.

BK：日本語書物

大泉充郎監修・藤村 靖編著：音声科学，東京大学出版会，1972 年 3 月.

中田和男：音声，コロナ社，昭和 52 年 10 月，

比企静男：音声情報処理，東京大学出版会，1973年 10 月.

(昭和 55 年 6 月 16 日受付)