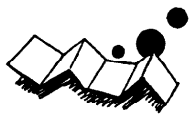
解　説



合成音声出力装置†

小池恒彦††

1. はしがき

1978年末，米国のテキサス・インスツルメント(TI)社は，わずか5mm角程度のシリコン・チップの上に音声合成器を実現し，これをSpeak & Spellと称する語学教育の玩具として米国内外で発売を開始した[1]．これは，単なる教育用玩具としての意味よりも，これまで進められてきた音声情報処理，大規模集積回路の研究のある意味での到達点として大きな意義をもつものであるといえる．

電子計算機の発達に伴って，計算機利用法の簡単化，すなわちマン・マシン・インタフェースの改善の努力が進められてきた．そのひとつとして，通常の電話機から，電話網を通して計算機にアクセスし，情報をやりとりすることがあげられる．これにより各種の情報案内，予約，科学技術計算などが，電話機操作のみで可能となり，きわめて便利となる．この場合，計算機からの出力手段として，人間の音声に近い信号を用いれば，出力情報の理解に訓練を必要とすることなく，容易に通信が可能である．こういった立場から，人間の音声に似た信号を機械的に出力できる装置の開発が始められた．約10数年以前のことである．

一方，音声の分析，合成，認識などいわゆる音声情報処理の研究は非常に古いが，機械的な出力音声へのニーズが出てきたのとちょうど時を同じくして，計算機を用いたディジタル信号処理技術を駆使した新しい音声分析合成技術が考案され，著しい発展をとげてきた．PARCOR分析合成方式などの，いわゆる線形予測符号化(LPC=Linear Predictive Coding)技術である．この技術は音声のスペクトルの特徴抽出に関するものであるが，応用面からみると，音声の低ビットレート伝送，認識，合成など広い範囲をカバーしている．

† Synthetic Voice Response Unit by Tsunehiko KOIKE (Musashino Electrical Communication Laboratory, N.T.T.).

†† 日本電信電話公社武蔵野電気通信研究所

このように合成音声のニーズとそれを支える音声合成技術，さらに近年著しく進歩している大規模集積技術が一体となって，TI社の玩具が実現されたといえよう．

本稿では，音声処理技術のうち，合成とその応用である音声出力装置について解説する．なお，音声の特徴抽出については，合成と密接に関係するが，本誌の解説[2]があるので，詳細はこれに譲るとして，概略の説明にとどめる．また，音声の全般的な解説については，参考文献を参照されたい[3)~5)]．

2. 音声出力装置の特質と要求条件

電子計算機の出力を音声で得るときの利用者側からみた利点，欠点を考えてみる．まず，利点をあげると次のようになる．

端末が簡易　キーボード・プリンタや，ディスプレイなど高価な端末装置を使用する必要がない．押しボタン電話機か通常の回転ダイヤル電話機に簡単に入力手段を付加したものでよい．

操作上の訓練が不要　計算機からの出力情報は，聞き馴れた音声メッセージであるので，特別な使用上の訓練はいらない．

同時性　出力情報を耳で聞きながら，その情報に関連する作業を同時に行うことができるため，作業能率の向上が期待できる．また，多数の人間に同時に，同じ情報が流せるといった利点もある．

半面，音声であるがゆえの欠点もある．

記録性　プリンタのように出力情報のコピーが残らないので，複雑な情報の出力にはむかない．反復出力が可能なようにすれば，この問題はある程度軽減される．

音韻の不明瞭性　合成音においては，各音韻の明瞭性が自然音声の明瞭性と異なる場合が起りうる．極端な例では，ある特定の音韻が合成できない場合もありうる．人間の会話では，不明瞭な場合，言いまわしを変えたり，何らかの冗長性を付加することで，明瞭性

を確保することができる．合成音声による通信では，これが困難であるので，各音韻の明瞭性を一定以上に維持する必要がある．

以上のような，音声出力装置の利点，欠点も，それを適用するサービスによって，重要性が異なるのは当然であるが，一般的には，端末装置におけるメリットを活かして，電話機を用いた簡易な情報案内，検索，予約，確認などに音声出力が利用されている．

次に音声出力装置が実用に供されるときに考慮される条件について考えてみると，第1は語彙の規模である．装置が出力すべき語彙はサービスによって異なる．ある場合は，数字や時刻，若干の定形的な表現などに必要な少数の単語を準備しておけばよいし，またある場合は，例えば人名など固有名詞を出力できる大規模な語彙を持つことが要求される．どの程度の語彙が必要かということは，音声出力装置でまず考えられねばならない点である．

第2は出力音声の品質である．音声品質の良さを表わす尺度には種々ある．肉声らしさ，すなわち自然性の確保も大切であるが，少なくとも，音声出力で何らかの言語的情報を伝達しようとする場合は，最低限了解性は保証される必要があろう．

3. 音声の生成

3.1 音声の波形とスペクトル

音声は，図-1に示すような発声器官によって生成される．すなわち，声帯の振動によって生じた空気流が舌，硬口蓋，軟口蓋，鼻腔，歯，唇などで形成される声道を通過するとき，声道の形状によって決まる特定の変調を受けた後，空気中へ放射される．声道の形状によって種々音韻の音響的特徴が定められる．また音声信号の中には，声帯振動を伴わないものもある．舌，

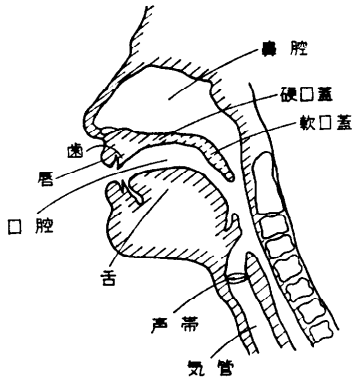


図-1 音声器官

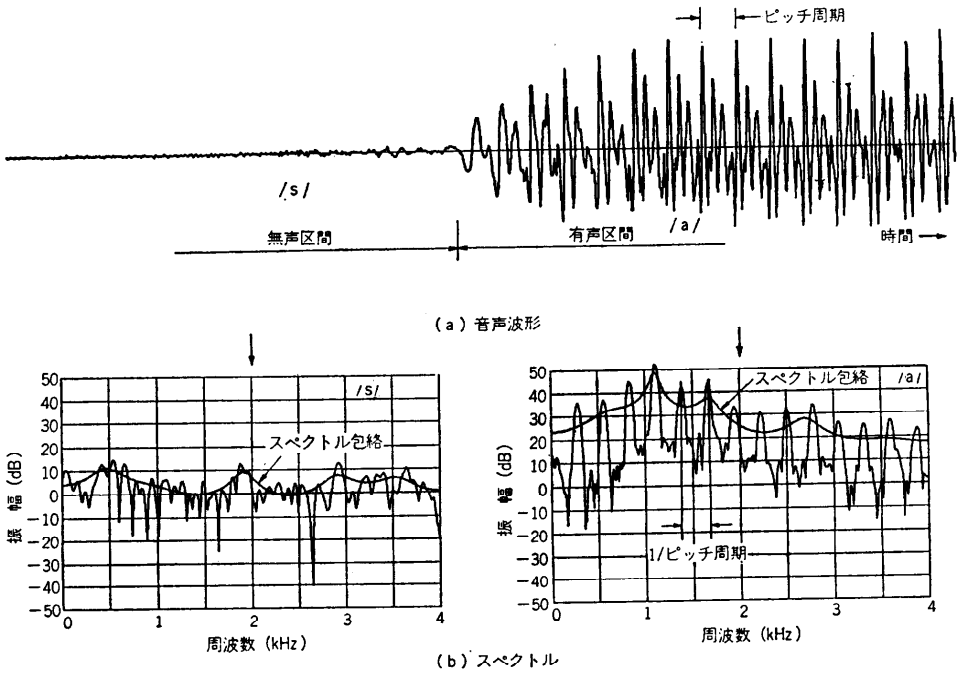


図-2 音声波形（a）とスペクトル（b）(女声/sa/の一部)

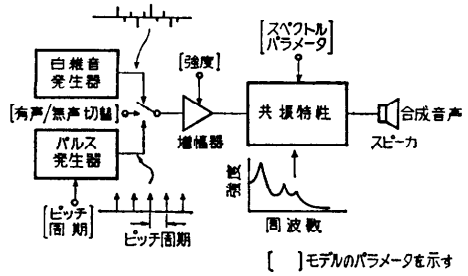


図-3 音声の合成モデル

歯，唇などにより声道の途中に作られた狭めを呼気が通過する際に生ずる空気の乱流や，声道の一部を閉じた後，急激に開放することにより生ずるインパルス性の空気流をもとにして音声が生成される場合である．

音声はその発生のメカニズムから細かく分類されているが，大きく分けて，声帯振動を伴うものを有声音，伴わないものを無声音と呼んでいる．

観測された音声波および周波数スペクトルの代表的な例を**図-2**に示す．これは女声の /sa/ の一部を拡大したものである．**図-2**(a)の左半分は /s/ の雑音的な波形であるが，右半分は /a/ でエネルギーの大きい，ほぼ周期的な波形のくり返しが見られる．/s/ では声帯が振動しないが，/a/ では声帯の振動があり，この振動の基本周期（ピッチと呼ぶ）が波形の上に明瞭に観察される．

また，/s/ および /a/ の一部を短時間スペクトル分析したものが，**図-2**(b)に示されている．/s/ の部分では，スペクトルの包絡にはいくつかの山が見られるが，微細な構造は非常に不規則な様相を示している．

一方，/a/ では，スペクトル包絡にはいくつかのピーク（ホルマントと呼ぶ）が観察されるとともに，微細構造として(ピッチ周期)$^{-1}$の調波構造が明瞭に存在する．

3.2 音声生成のモデル

人間の発声器官による音声の生成は，前項で述べたように，大ざっぱに言って，声帯波または声道の一部から生ずる乱流によって，声道という音響管が駆動され，その出力が空気中に放射される過程といえる．すなわち，駆動音源のスペクトルを $G(\omega)$，声道の伝達特性を $V(\omega)$，唇からの放射特性を $R(\omega)$ とすれば，生成される音声のスペクトル $S(\omega)$ は次式で近似的に表わされる．

$$S(\omega)=G(\omega)V(\omega)R(\omega) \qquad (1)$$

有声音では $G(\omega)$ が調波構造をもつので，図-2(b)の /a/ のスペクトルのように，包絡がほぼ $V(\omega)$ で決まる調波構造をもつスペクトルが得られ，無声音では包絡が $V(\omega)$ の連続スペクトルが得られる．

(1)式のモデルは次のようにも考えられる．$S(\omega)$ のスペクトル包絡特性をもつ回路を，平坦なスペクトルの音源で駆動すれば，スペクトル包絡特性は再現できる．さらに，ピッチ構造を付加するために，有声音の場合にはピッチ周期ごとのインパルス列を，また無声音の場合には白雑音を音源とすればスペクトルの微細構造まで再現できることになる．この考えによる音声の合成モデルを**図-3**に示す．

図-3のモデルに現われるパラメータとしては，$S(\omega)$ の包絡特性を記述するスペクトル・パラメータ，ピッチ周期，有声／無声の識別情報，音声の強度の4種類が必要である．

この中で $S(\omega)$ の包絡特性の記述自体と，音声のように時間的に変化する動的特性をどのように表現するかが問題であるが，通常，動的特性については，声道の変化の速度が比較的ゆるやかであることから，20 ms前後の短時間の間は特性は不変であると考える．また前者のスペクトル特性の記述に関しては，聴覚的に位相特性の寄与が第2義的であるものと仮定して，その振幅特性だけを記述するのが通例である．

振幅特性の記述の方法は種々研究されている．その代表的なものとして，古くは，帯域フィルタ・バンクによる分析，短時間自己相関分析，ホルマント分析などから，最近では PARCOR（偏自己相関）分析に代表される LPC（線形予測分析）などがある[2]．これらの分析により得られたパラメータを図-3のモデルに10～20 ms ごとに供給すれば，音声が合成される．

図-3の合成モデルは周波数領域で表現されているが，時間領域でも同じように考えられる．すなわち $S(\omega)$ の包絡特性の代りに，そのフーリエ変換であるインパルス応答波形を求め，これを駆動音源からの入力パルス列に同期して接続していくことにより，合成音声が得られる．

4. 音声出力方式

4.1 音声出力方式の分類

3章において音声合成の工学的モデルに関して述べてきたが，電子計算機の処理結果を音声の形で出力するといった実際的な目的のためには，必ずしも合成技術を使わなくとも録音・再生技術で十分対処できる場合がある．これらを含めて，音声出力方式の分類を行うと**図-4**のようになる．この分類は，音声信号の波形，

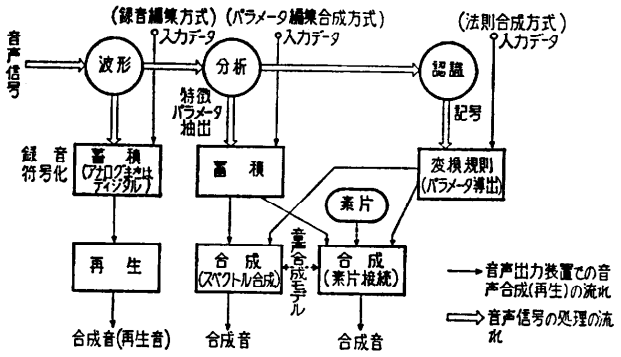


図-4　音声出力方式の分類

特徴パラメータおよび言語的単位としての記号（音韻，音節など），といった3つの音声記述レベルとの相互関係で，音声出力方式を分類したものである．

図-4に示されるように，音声出力方式は通常，大きく分けて，録音編集方式，パラメータ編集合成方式および法則合成（またはルール合成）の3方式に分類される．

各方式の特徴の比較を表-1に示す．比較の要因としては，2章で述べた実用上の条件である品質および語彙，さらに装置経済性に影響を与える複数出力を同時に得る多重化制御の3要因を選んである．

また，各方式で合成音（品質的には同等とはいえないが）を得るのに必要な情報量の概略の値を表-2に示した．

以下に各方式について説明する．

4.2　録音編集方式

波形レベルの録音・再生で音声出力を得る方式である．あらかじめ，出力に必要な単語や文節を録音しておき，これらを選択的に再生して，つなぎ合わせることにより，音声メッセージを組み立てる．実用的には，録音媒体として，アクセス時間の関係から，アナログ録音形の磁気ドラムやトーキ録音の原理による光学ドラムが用いられる．ドラムの各録音トラックに，単語などを録音しておき，ドラムの回転に同期してトラックを切替えつつ再生していけばメッセージが組み立てられる．また，波形をPCMなどの符号化処理してディジタル磁気ドラムに蓄積することもできる．

この方式は，自然音声の録音・再生が原理であるので，単語や文節の接続点で生ずる若干の不自然さを除き，全体としては品質は良好である．半面，アナログ磁気ドラムなどの容量から録音できる単語や文節の数は100前後にすぎない．大形の磁気ドラムにディジタル録音する場合でも，PCMの音声は64 Kbit/secであり，1秒の単語の数に換算して約500語が限度である．

また，多数出力を同時に得る多重化制御は，全録音トラックを常時読み出し，時分割スイッチを用いて出力端子へ切替えることで容易に可能である．

4.3　パラメータ編集合成方式

録音編集方式における語彙の拡大をねらったのが，パラメータ編集合成方式である．したがって，音声メッセージを組み立てる方法は，録音編集方式と変らない．異なる点はソフト的には音声が特徴パラメータの形で蓄積されること，ハード的には再生機構の代りに音声合成器が用いられることの2点である．

本方式には，図-3の合成モデルを周波数領域で実現するか時間領域で実現するかによりスペクトル合成，素片接続合成の2種類がある．スペクトル合成では，(1)式における $S(\omega)$ の包絡特性を記述する特徴パラメータから，元の包絡特性を再現するようなフィルタが必要となり，一方素片接続合成では，フィルタの代りに，素片波形を蓄積し，これを必要に応じて読み出し，接続する処理が必要となる．

音声品質の点では，録音編集方式にくらべて，情報量が1/10～1/30まで削減されているので，全体的には劣化がある．特に，特徴パラメータとして，帯域フィルタ・バンクの出力や短時間自己相関係数などを用

表-1　音声出力方式の特徴

項目 / 方式	音声品質	語彙	多重化制御	備考
録音編集方式（アナログ/ディジタル）	了解性：高 自然性：高	小（500語程度以下）	容易	実用化済
パラメータ編集合成方式	了解性：高 自然性：中	大（数千語）	多重合成または LSI 合成器	初期の実用化段階
法則合成方式	了解性：中 自然性：低	無限	法則の複雑さに依存する	法則に関する研究を要する

表-2　合成に必要な情報量の比較

録音編集	PCM	64 Kbit/sec
	ADM	25 Kbit/sec
分析合成	PARCOR	2,400 bit/sec
	ホルマント	数 100 bit/sec
法則合成		数 10 bit/sec

いると劣化が目立つが，最近急速に進歩した LPC 系のパラメータ，例えば PARCOR 係数，線形予測係数などを用いると，かなり高品質な合成音が得られる．品質については後述するが，パラメータ編集合成方式では了解性はほとんど録音編集方式並み，自然性にやや改善の余地があるというのが現状である．

語彙については，大形の記憶装置を用いれば，数千単語まで拡大することができる．これを逆に見れば，録音編集方式で実現される程度の語彙の装置はパラメータ編集合成方式を用いれば非常に小形化できるということになる．

合成器のハードは，スペクトル合成では LPC 系のパラメータを用いたディジタル・フィルタで，$S(\omega)$ の包絡を実現するのが普通である．ハード構成は乗算器，加算器，遅延用のレジスタ，その他駆動音源やパラメータ補間などの論理回路を含めて 4～5 千ゲートの論理規模である．合成に先だち，パラメータの非線形変換が行われることもあり，このための変換テーブルとして ROM が加わる場合もある．

このように合成器を用いる場合は，出力の多重化は，時分割多重合成器を用いるか，空間分割的に合成器制御するかによって行われることになる．

また，素片接続合成では，音声素片を蓄積するメモリの他は振幅制御用の簡単な論理で十分であり，多重化制御も比較的簡単である．なお，音声素片としては，自然音声から抽出したインパルス応答波形，自然音声を LPC 分析合成して得られたインパルス応答波形，声道模型から求めたインパルス応答，ホルマントに対応した減衰正弦波などが，実用的な見地から考案されている[5]．

4.4 法則合成方式

音声信号は音韻などの離散的な言語記号が，人間の音声発生過程によって連続的に波形に変換されたものとみなすことができる．法則合成方式はこのような変換を機械的に実現しようとするものである．したがって，入力は音韻などの記号系列が与えられる．このことは，どんな単語，文章でも，記号系列で表現されたものは音声に変換できることを意味する．すなわち語彙は無限である．しかしながら，実際には，自然な肉声に近い合成音を得るには困難な問題が多い．問題は図-4 における変換規則である．パラメータ編集合成方式では生の音声をもとにしているので，できるだけこれを忠実に再現するようなパラメータの設定を行えばよいが，法則合成では規則により作り出さねばならない．困難さのひとつは，自然音声では音韻間の相互作用による音韻特徴パラメータの変形がみられることに起因する．これは調音結合と呼ばれる現象であるが，この変形の規則の良否が法則合成音の品質に直接反映する．

法則合成においても，記号に対応して基本的な音声の単位は準備される．日本語の場合には，これらの単位として，音韻（音素），音節，VCV 連鎖（母音・子音・母子連鎖）[6]などが用いられている．VCV は，母音と子音が交互に表われるという日本語音声の特徴を活用した単位であり，調音結合の影響を受けやすい子音を前後の母音ではさんでいるので，単に VCV 連鎖を結合するだけでも，比較的自然な合成音が得られる．日本語の場合，音韻の数は数十，音節の数は 100，VCV では約 770 個必要といわれている．

法則合成を用いた音声出力方式は，固有名詞のようにほぼ無限の語彙が要求される場合や，電子計算機利用者が自ら出力メッセージを設定できるようなシステムでは不可欠であるが，現在のところ，十分な品質を与える規則はまだ研究段階といえる．

5. 音声出力装置の実際例

5.1 音声出力装置を用いたシステム構成

図-5 は押しボタン電話機を端末機とし，交換機を通して，中央の処理装置にアクセスし，結果を音声で返答するシステムである．音声合成部では，中央処理装置からの出力指令にもとづき，音声メモリに蓄積されたデータをもとに，音声合成部で音声を合成（または編集）しつつ，電話機へ返送する．

このように，システムの基本的要素としては，中央処理装置，音声出力装置，端末とのインタフェース部がある．

現在，実用に供されているサービスの一覧表を表-3

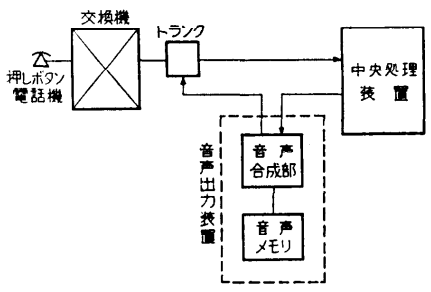


図-5 音声出力装置を用いたシステム構成

表-3 音声出力サービスの実際例

音声出力方式		会社等(実施年)	サービス名	音声出力装置の仕様			備考
				語彙	回線数	音声メモリ	
録音編集方式*	アナログ録音形	公社(45年)	電話計算サービス	128	512	アナログドラム	
		公社(47年)	時報サービス	39	1	同上	
		富士通(54年)	勝馬投票券電話投票サービス	512(128×4)	128	光学フィルムドラム	
	ディジタル録音形	松下(46年)	札幌高速鉄道自動案内	38	51	ディジタルドラム	
		富士通(46年)	成田空港自動案内放送	584	4	同上	
		日立(50年)	国鉄座席予約サービス	2,048	768	同上	
分析合成方式(LPCパラメータ)		公社(55年)	展示システム(PARCOR形音声応答装置)	2,048	128	磁気バブル	
		TI社(53年)	Speak & Spell	～200	1	ROM	教育玩具
法則合成方式**(音節編集PCM単)		日立(54年)	富士銀行電話連絡サービス	2,048+任意語	128	—	振込通知残高照会
		日電(予定)	住友銀行テレバンキングサービス	2,048+任意語	192	—	同上

* 録音編集方式の米国の実施例は省略した.
** この他，装置単体として米国 Federal Screw Work 社製の VOTRAX が発売されている.

に示す．以下，代表的なサービスとその音声出力装置の特徴について述べる．なお，表-3において，法則合成方式に分類されているサービスは，日本語音節を単に接続する音声出力方式を用いているので，法則合成とはいえないが，無限語彙が出力できる点から便宜上このように分類しておく．

5.2 国鉄座席予約サービス

押しボタン電話機から列車名，発着駅などの情報を送り，音声応答装置と会話しながら，予約を完成させるサービスである．全国の特急停車駅の駅名をすべて音声出力するために，語彙が2,000語以上となっている．

出力方式はディジタル録音編集方式であり，音声メモリとしては，大容量の磁気ドラムを複数台用いている．音声はRCM符号化されたもので品質は良好である．1台のドラムの容量は500単語が限度であるが，この音声出力装置では，複数台のドラムとバッファメモリを用いた同期技術により2,000の語彙が実現されている．

5.3 PARCOR 形音声応答装置[8]

電電公社において，ユーザ向けの展示システムに使用される装置である．9.6 Kbit/s の PARCOR 係数を用いたパラメータ編集合成方式であり，品質的にはかなり良好である．図-6 に PARCOR 形音声応答装置のブロック図を示す．この図では音声メモリとして磁気ドラムが用いられているが，展示システムでは，磁気バルブが用いられている．合成器は8多重合成器を4個まとめて32多重としたものをユニットとして，増設することができる．合成器の論理的な構成を 図-7 に示す．係数の次数 P は10である．ハード的には加算器ですべての演算が実行されている．

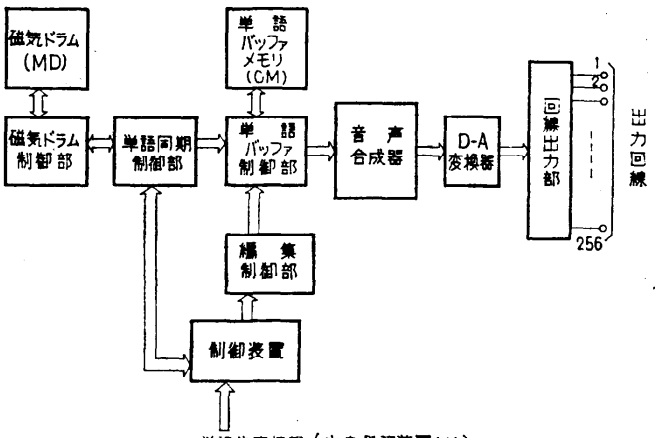


図-6 PARCOR 形音声応答装置の構成

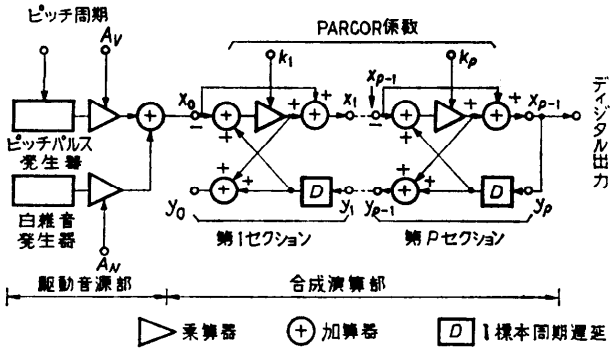


図-7 PARCOR 形音声合成器の構成

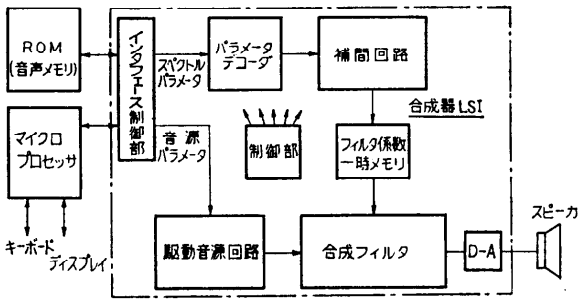


図-8 Speak and Spell の構成

5.4 Speak & Spell[1)]

英語のつづりの練習や，つづり当てゲームができる一種の玩具である．この装置は図-5に示すシステム構成の中で交換機を除くすべての構成要素を小さな箱の中におさめたものである．押しボタン電話機はキーボードとスピーカで，中央処理装置はマイクロプロセッサにより，また音声合成部は p-MOS の 1 チップ LSI で，音声メモリは 16 K Byte の ROM 2個で，それぞれ置きかえられている．

音声出力方式はパラメータ編集合成方式に分類されるもので，LPC 分析された音声パラメータが符号化されて ROM に蓄積される．パラメータは 1 組 48 bit が 20 ms ごとに合成器に供給されるが，1 回前に与えられたパラメータと次回に与えるべきパラメータの間に差が小さい場合は，前回のパラメータを使うようにしてパラメータの情報レートの削減をしている．この結果，平均的には 2 個の 16 K Byte ROM に 200 秒の音声を収容できるようになっている．

Speak & Spell の構成を図-8に示す．全体の制御はマイクロプロセッサが行っている．音声出力時は，マイクロプロセッサから音声の先頭アドレスが合成器に与えられ，これにより，合成器はパラメータを ROM から読み出し，音声を合成する．合成器に取り込まれたパラメータのうち，スペクトル・パラメータはデコードされた後，2.5 ms ごとに補間されて合成用のフィルタ係数とされる．また音源パラメータは駆動音源部に供給されて，フィルタの入力パルスが発生される．合成器の主要部である合成フィルタは，等価的に図-7 に示すフィルタと同じ演算を行うもので，ハード的には 14 bit の加算器と 10 bit×14 bit の乗算器各 1 個と，遅延用のいくつかのシフトレジスタから構成されている．

合成器チップ上には，合成フィルタはもちろん，合成に必要なすべての回路が実現されている．特徴的なことはチップ上に組み込まれた D-A 変換器 (8 bit) が直接スピーカを駆動するように設計されていることである．

なお，TI 社の合成器 LSI と同様の LSI が昭和 54 年電電公社と日立の共同で開発されている．

5.5 Votrax 合成器[9)]

これは図-4 に示された法則合成方式に分類される音声出力装置で，キーボードまたは電子計算機などから与えられた離散的な記号——音韻記号，ピッチ，音声の速さ——から変換法則によって合成に必要なパラメータ系列を作り出す．合成はスペクトル合成方式であり，5 個の共振回路と 1 個の反共振回路で声道の伝達特性を表現している．

本装置の構成を図-9に示す．1 音韻あたり 12 bit のデータが入力される．このうち 7 bit は音韻の種類を指定し，5 bit が韻律（うち 3 bit がピッチ，2 bit が早さ）を指定する．これらの入力データは発声・調音制御部とフラグ制御部においてパラメータに変換された後，時間的に変化するアナログ制御信号として音声合成部に供給される．音声の高さや早さ，音量は手動設定によっても変えることができる．

6. 合成音声の品質

録音編集方式による出力音声はともかくとして，パラメータ編集合成や法則合成の出力音声は，やっと身近かなものとなり始めたというのが本当のところであろう．したがって合成音声の品質に対して，一般的な考え方が確立しているわけではないが，ここでは合成音声を用いた通信を音声通信とみなして，品質につい

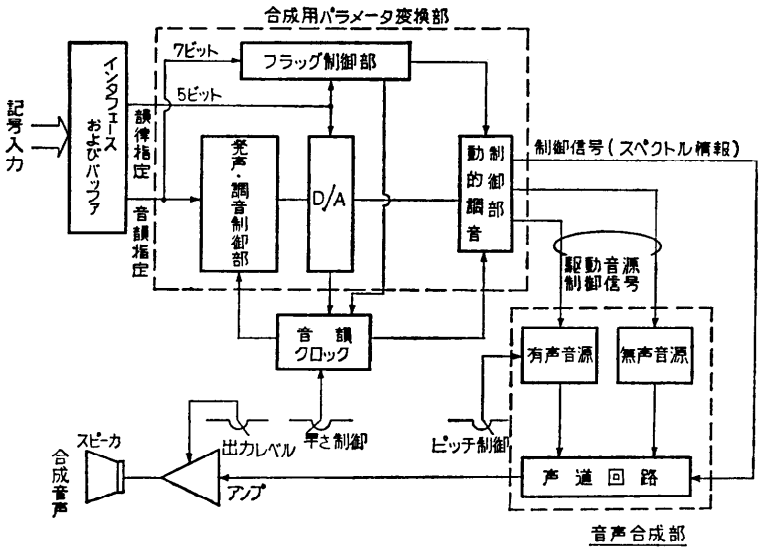


図-9　Votrax-ML 1 の構成

て考察してみる．

6.1 品質尺度

電話伝送系の品質評価尺度として古くから用いられてきたものとして，単音明瞭度がある．この尺度は，日本語 100 音節で受聴テストを行い，子音と母音を分離して集計した平均的な明瞭度である．この尺度と文章了解度の関係は自然音声の場合について研究されている．すなわち，単音明瞭度が 80% あると 50% の人の文章了解度が 100% になることが明らかにされているので，合成音のうちに言語内容の通信を目的とする場合の了解性の評価尺度には適している．

次に合成音声の特殊性として，個別的な音韻の明瞭性が自然音声と比較して著しく異なることも考えられる．極端なことを言えば，ある特定の音韻が出力できないような場合も起り得る．単音明瞭度は個別音韻の明瞭度から求めた平均的な明瞭度であり，個別音韻の明瞭度の分布には無関係である．したがって，合成音では，単音明瞭度だけでは了解度が必ずしも保証されるとはいえない．個別音韻の評価尺度として正聴率が考えられる．正聴率は各音韻の誤りにくさを表わす尺度で，明瞭度テストで求められる．

データ通信の出力として，数字が多く用いられるが，数字は冗長性が少ない上に，誤まると影響が大きいため，品質尺度として，特に数字の了解度を評価することも考えられる．

了解性に着目した尺度以外に，肉声らしさ，すなわち自然性に関する尺度も考える必要があろう．電話通話においては，オピニオン評価（通話の良さを段階評価する）により通話の総体的な評価を行っているが，合成音の自然性評価にも，類似の評価が考えられる．

6.2 品質評価

前項で種々な評価尺度をあげたが，音声通信で最も重要である了解性を評価する単音明瞭度にもとづいて，各種の方式を評価してみる．

ここで AEN（明瞭度等価減衰量）という量を導入する．単音明瞭度そのものも品質尺度であるが，受聴テストにおいては試験員によりバラツキが生ずるた

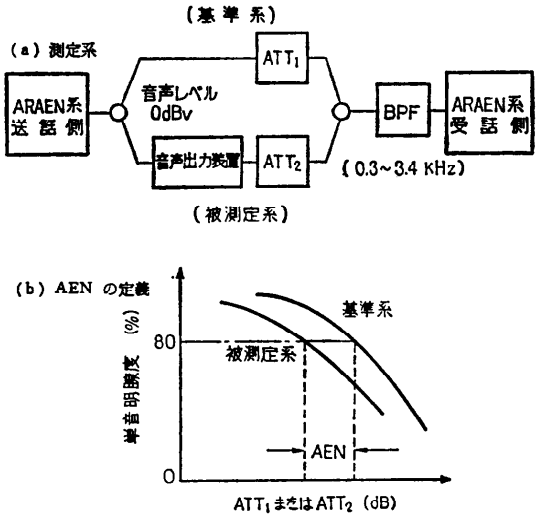


図-10　AEN の測定法

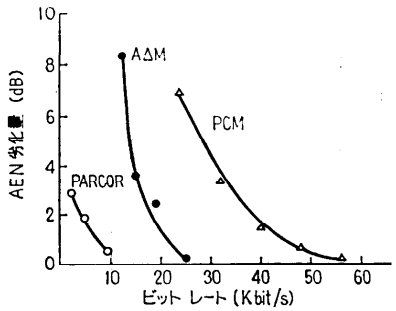


図-11 各種出力方式の AEN 劣化量

め，基準となる系を定め，その系と被試験系との相対的な量として AEN を定義する．**図-10** に AEN の測定法を示す．(a)の測定系において基準系は，2人の人間が1m離れて対面したときの空間の音響特性を電気的に実現した系（ARAEN 系と呼ばれる）に帯域制限を加えたものである．基準系，被測定系で，それぞれ減衰量を変えて単音明瞭度を測定する．図-2(b)ような曲線を描き，単音明瞭度 80% の両系の減衰量の差で AEN が求められる．

このような方法で，PARCOR 係数を用いたパラメータ編集合成方式，および ADM（適応形 △M），PCM の品質評価を行った結果を**図-11** に示す[11]．この図に示すとおり，それぞれの方式の適用領域が明確に分かれていることがわかる．PCM や ADM のような波形符号化はビットレートを上げると明瞭度は向上していくが，PARCOR のようなスペクトルを符号化する方式は，10 Kbit/s 程度で飽和する傾向がある．

AEN の劣化量をどこまで許容するかは，音声出力装置が使われるサービスによるが，電話網の端末一端末間で合成音声が使われる場合に限れば，電話網の限界的な接続状態でも 80% の明瞭度を保証する必要があり，AEN 劣化量として数 dB になることが考えられる．

7. あ と が き

音声の合成とその応用である音声出力装置について解説した．最近の研究である合成音品質評価についても若干ふれた．合成音声が一般的に許容される品質の水準は，今後合成音声が社会生活に浸透していく過程で次第に決まっていくものであろう．

いずれにしても，Speak & Spell の各方面へのインパクトは大きい．1980年代は，あらゆる分野で合成音声が手軽に使われることとなろう．通信網，電子計算機，端末機器などが口を持つことにより，電気通信のマン・マシン・インタフェースが改善されることはもちろん，家電製品，自動車，交通機関，大小プラントなどで合成音による機能向上がはかられるという形で．

参 考 文 献

1) Wiggins, R. and Brantingham, L.: Three-chip system synthesizes human speech, Electronics (Aug. 31, 1978).
2) 板倉，東倉：音声の特徴抽出と情報圧縮，情報処理，Vol. 19, No. 7, pp. 644-656 (1978).
3) Flanagan, J. L.: Speech Analysis, Synthesis and Perception, 2nd edition, Springer-Verlay (1972).
4) 大泉充郎監修，藤村 靖編：音声科学，東大出版会(1972).
5) 中田和男：音声，日本音響学会編音響工学講座7 (1977).
6) 斉藤，橋本他：音韻連鎖に着目した音声合成システム，日本音響学会研究発表会，3-2-3 (May 1963).
7) Sagawa, S. et al.: Automatic Seat Reservation by Touch-Tone Telephone, 2nd USA-Japan Computer Conference Proceedings (Aug. 1975).
8) 小池，木下他：PARCOR 形音声応答装置，研究実用化報告，Vol. 23, No. 10, pp. 2107-2120 (1974).
9) Sherwood, B. A.: The computer speaks, IEEE Spectrum (Aug. 1979).
10) 電子通信学会編：聴覚と音声 (1966).
11) 北脇，伊藤他：音声応答装置の品質評価に関する考察，日本音響学会，研究発表会，S 79-42 (1979-10).

(昭和54年12月17日受付)