U.D.C. [534.782:681.323]:[654.922.8:656.52.072.7:621.876]

音声合成技術の現状とその応用

——エレベーター・エスカレーターへの応用——

Speech Synthesizing Technology and Its Application —Application to Elevator and Escalator—

弓仲武雄* *Takeo Yuminaka*
三瓶　徹** *Tôru Sampei*
野宮紘靖*** *Hiroyasu Nomiya*
中田和男**** *Kazuo Nakata*

音声合成装置は機械から人間への警報，案内，指示などを出力し，人間と機械のインタフェースを容易かつ確実にするもので，通信，コンピュータ，産業機器など広い用途をもっている。しかし，従来は音声合成技術が一般に普及していないこと，回路規模が膨大で高価なことから，ごく限られた用途にしか実用化ができなかった。この課題に対し，LSI技術によって1チップ化，低価格化を図り，性能的にも産業用，コンピュータ用にも活用できる高性能をねらい，音声合成の特長を生かした機能も開発して用途の拡大を図った。

本稿では，音声合成方式の原理とシステム構成の概要について述べ，その応用分野と代表的な実施例として，エレベーターへの応用について紹介する。

1 緒　言

音声合成装置は，機械から人間へのコミュニケーション，すなわち警報，案内，指示などを音声によって瞬時に確実かつ安定して出力できるという特長をもっており，通信，コンピュータ，産業機械，自動車，時計，教育機器，家電品などの機能向上に広い用途をもっている。しかし，従来は音声合成技術が一般に普及していなかったこと，回路規模が膨大で高価なことから，ごく限られた用途にしか実用化されていなかった。

日立製作所では，その将来性に注目し，音声合成部の低価格化，高性能化及び応用の拡大を目的として音声合成LSIの開発を計画し，昭和53年秋から研究を開始した。

音声合成法としては，日本電信電話公社で発明され，現在世界の主流となっているPARCOR(Partial Autocorrelation：偏自己相関)法[1]を採用し，日本電信電話公社の適切な指導により，昭和54年9月に国内で初めて音声合成LSIの開発に成功し，昭和55年2月から量産に入り，音声端末装置，エレベーターの音声案内，自動車警報装置，音声時計付ラジオ，珠算の読上算練習器などに実用化し，更に現在広く産業用，通信用，コンピュータ用から家庭電気品用に至るまで広い分野で採用されつつある。

以下，音声合成法，開発したLSI及び応用の一例としてエレベーターの自動放送装置につき紹介する。

2 音声合成の方式

PCM(パルス符号変調)は，ディジタル通信の代表的な方式である。しかし，1秒の音声を再生するのに64kビット程度のデータ量を必要とするので，音声合成にはあまり使われない。そこで，データ量を圧縮するために音声信号の定常性などを考慮して，能率の良い符号化が行なわれる。音声サンプル値と前のサンプル値との差を送ることにより，量子化ビット数を減らしたDPCM(Differential PCM：差分PCM)，振幅の大きいところでは量子化幅を大きくして信号変化に追随するようにし，振幅の小さいところでは量子化幅を小さくし，小さな信号変化を再現できるようにしたAPCM(Adaptive PCM：適応PCM)やADPCM(Adaptive DPCM：適応差分PCM)などが一般的である。しかし，いずれの方式も符号化に工夫はあるものの結局は波形を伝送したり，記憶再生をしているにすぎない。

一方，音声信号の波形ではなく，原音声から音声を特徴づける幾つかのパラメータを抽出しておき，そのパラメータから音声を合成すると，大幅にデータ量を小さくすることができる。LPC[2](Linear Predictive Coding：線形予測符号化)，PARCOR,LSP[3](Line Spectrum Pairs：線スペクトル対)やCSM[4](Composite Sinusoidal Modeling：複合正弦波)と呼ばれる方法がそれである。

更に，データ量を小さくする方法として規則合成が挙げられる。先に挙げた方法が原音声を手本にして音を合成するのに対し，規則合成では手本なしに規則だけで音声を合成するので自然な韻律を与えることが難しい。

図1は，各方式を必要データ量(ビット/秒)軸上に並べたものである。音声合成の性能を決めているのは，音質とデータ量である。PCMなどの波形符号化法はデータ量さえ増せば音質は良くなる反面，データ圧縮が難しい。PARCORなどのパラメータ合成法は，データ量を増しても音質には限界があるものの，2.4kビット/秒というデータ伝送の情報量程度まで圧縮しても，実用に耐えられる音質が得られる。

日立製作所では，合成音の品質が良いこと，分析合成が完全な逆操作として定量的に対応でき計算処理が行ないやすいこと，昭和47年ごろから日本電信電話公社通信研究所の技術指導を受けたことなどの点から，音声合成法としてPARCOR方式を採用した。

3 PARCOR音声合成システム

PARCOR音声合成法で用いられるパラメータは，PARCOR係数である。PARCOR係数は，数学的には音声波形の偏自己

* 日立製作所水戸工場　** 日立製作所家電研究所　*** 日立製作所武蔵工場　**** 日立製作所中央研究所 工学博士

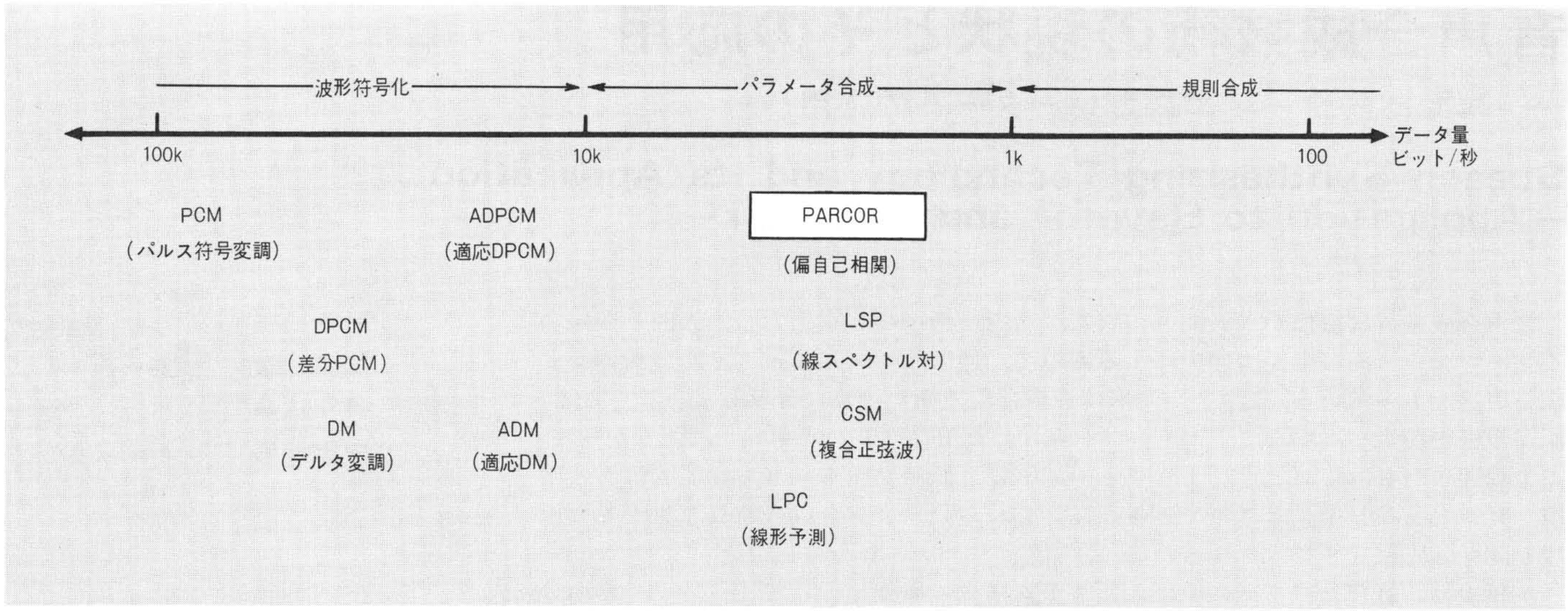

図1 音声合成方法と必要データ量 波形符号化法とパラメータ合成法の境界は10kビット/秒である。規則合成で使用するパラメータもPARCORやLSPであることが多い。

相関係数であるが，物理的には声道(声帯からくちびるまでの範囲)を多段の円筒管でモデル化したときの反射係数で，音声合成の過程は実際の人間の発声機構によく似ている。

音声は**図2**に示す声帯，口こう，舌，くちびるなどの器官によって作られる。まず声帯が持続的な振動を起こすと，呼気流は脈動的に振動する。この脈動の周期は声の高さを決定しており，男声の場合100～150Hz，女声の場合が250～300Hzで，この呼気流が口こう，くちびるで固有の共鳴特性が与えられ，音声となって放射される。

図3はPARCOR音声合成法での声道モデルを導く過程を示している。声帯の部分からくちびるに至る声道は，長さ15～17cmの太さの変化する円筒管としてモデル化できる。円筒管の接続部では，インピーダンスの不整合があるので音が反射し，共鳴特性が与えられる。これを電気回路に置換したのが同図(b)で，音波の流れを進行波と後退波とに分け，それぞれからK_iなる反射係数で反対側の波に加えている。これらの演算はすべてディジタル値の乗算，加減算の組合せで行なわれる。あらかじめコンピュータなどで分析して，メモリに収納しておいたPARCOR係数(声道をモデル化したときの反射係数)，音量，音源の種類と声の高さの情報を合成器に加えて合成する。声道は緩やかにしか変化しないので，PARCOR係数は一定の間隔(フレーム)で与えればよく，一般に10～30ms

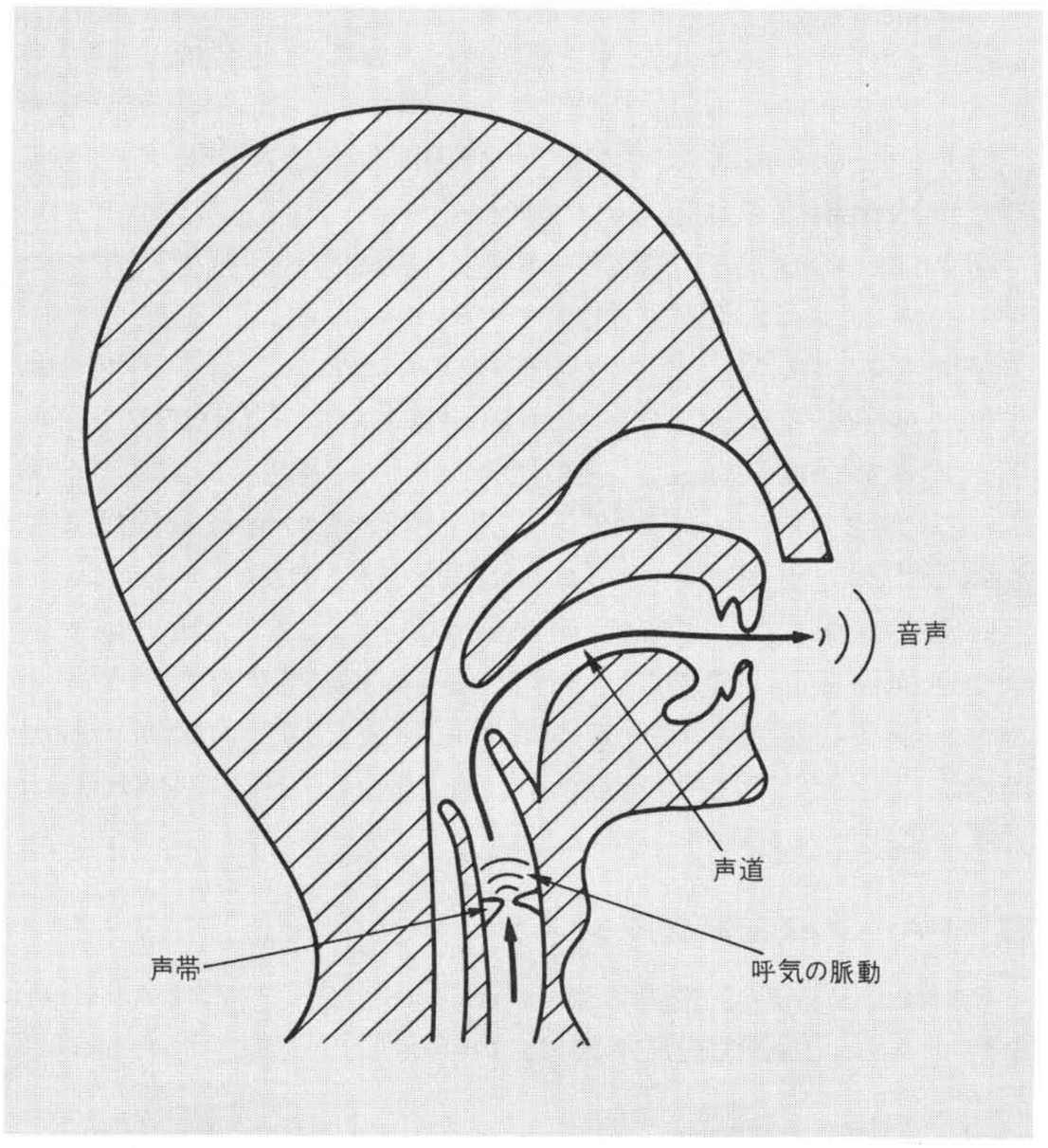

図2 人間の発声機構 声帯で引き起こされた呼気の脈動は，声道の共鳴特性を与えられ，口から音声として放射される。

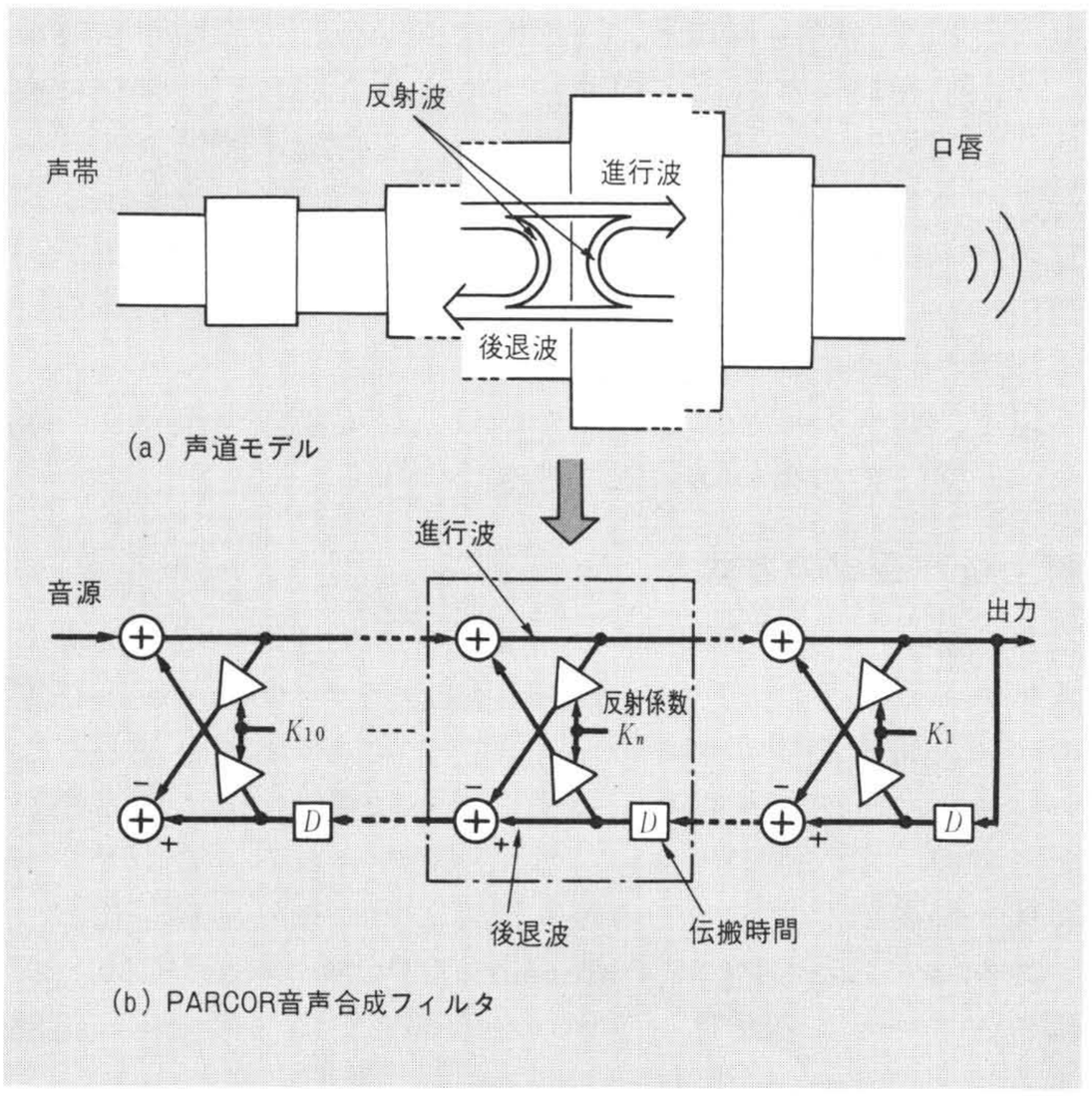

図3 声道モデルとPARCOR音声合成フィルタ 声道は，太さの変化する円筒管としてモデル化できる。円筒管の接続部でのインピーダンスの不整合により音が反射し，共鳴特性を与える。これを，電気的に(b)図のように乗算器，加減算器で構成する。

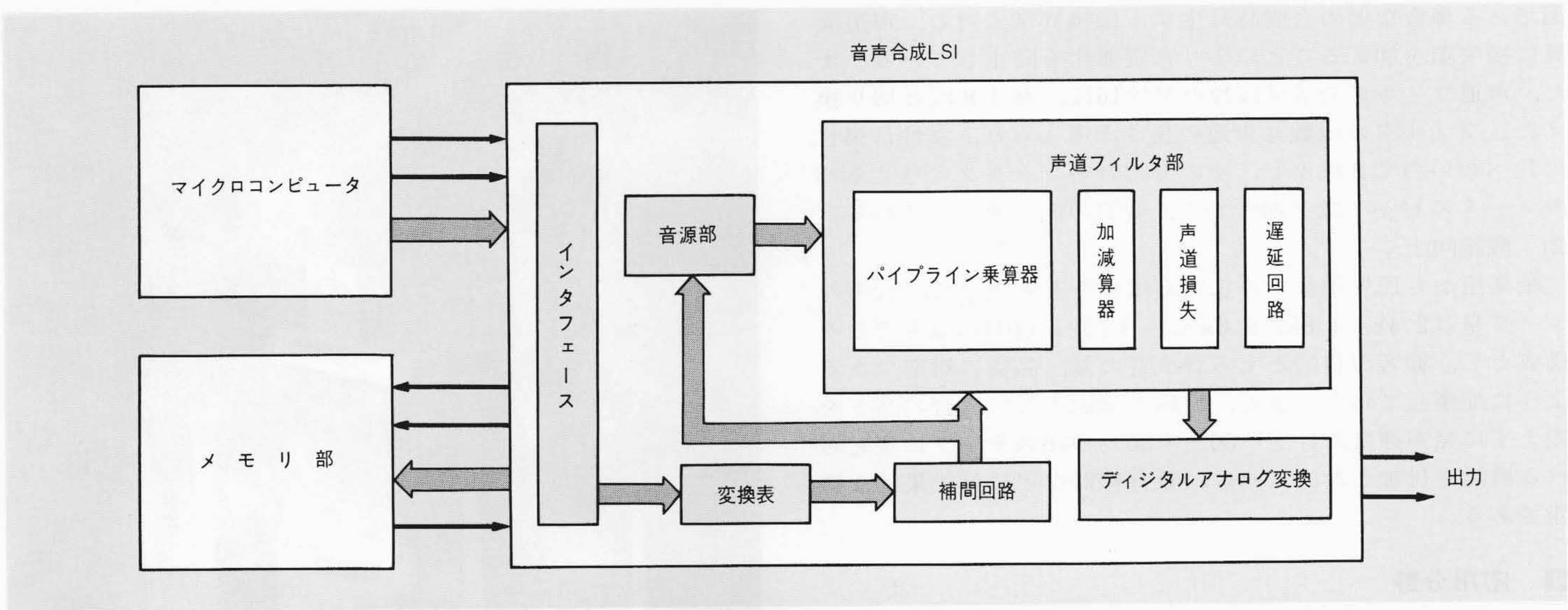

図4　音声合成システムブロック図　マイクロコンピュータは汎用の4ビット又は8ビットが，またメモリは大量生産用に専用マスクROM（128kビット）が用意されている。

が選ばれる。1フレーム分のデータはPARCOR係数，音源情報合わせて50～100ビットで済むので，1秒間の音声は1k～10kビットで合成でき，単純なPCMに比べ大幅にメモリを節約できる。

日立製作所の音声合成システムの仕様とブロック図，音声合成LSIのチップ写真をそれぞれ**表1**，**図4**，**5**に示す。システムは，(1)あらかじめ原音声からコンピュータで分析抽出したパラメータを記憶するメモリ部，(2)パラメータから音声を合成する音声合成LSI，(3)システムの動作を制御するマイクロコンピュータから成る。以下，本システムの特長について述べる。

表1　音声合成LSI仕様　声道損失演算の採用や，男性声音と女性声音で声道フィルタの段数を切り換えることにより，特に女性声音の品質が改善されている。

項　　目	仕　　様
合成方法	PARCOR（偏自己相関）
データ量	(1)　2.4kビット/秒　　大容量向け (2)　4.8kビット/秒 (3)　9.6kビット/秒　　高品質
発声速度	−60～＋30%　10ステップ可変
声道フィルタ	(1)　男性声音10段，女性声音8段 (2)　声道損失演算 (3)　演算精度　15ビット
出力部	(1)　ディジタルアナログ変換　精度±8ビット (2)　ディジタル出力15ビット (3)　スピーカ直接駆動
音源部	(1)　有声音はパルス又は$\mathrm{Sin}^2\,\omega t$波 (2)　無声音は13ビットM系列 (3)　外部信号で声の高さ制御可

(1)　合成音の品質向上

十分な品質の合成音を得るために，音声合成LSIの演算精度は15ビットに設定した。演算ビット数を増加すれば，合成音の品質向上が図れるのは当然であるが，回路規模から制限を受ける。コンピュータシミュレーションの結果，合成音の品質向上が頭打ちになるビット数は15ビットであることが分かり，品質と経済性を考慮して決定した。

また，従来の音声合成技術は一般に女性の声の合成を苦手としてきた。しかし，我が国の市場では民生用，産業用を問わず音声による情報のサービスには女性の声が重要視される。そこで，本音声合成LSIでは，女性の声のうち音質悪化の原

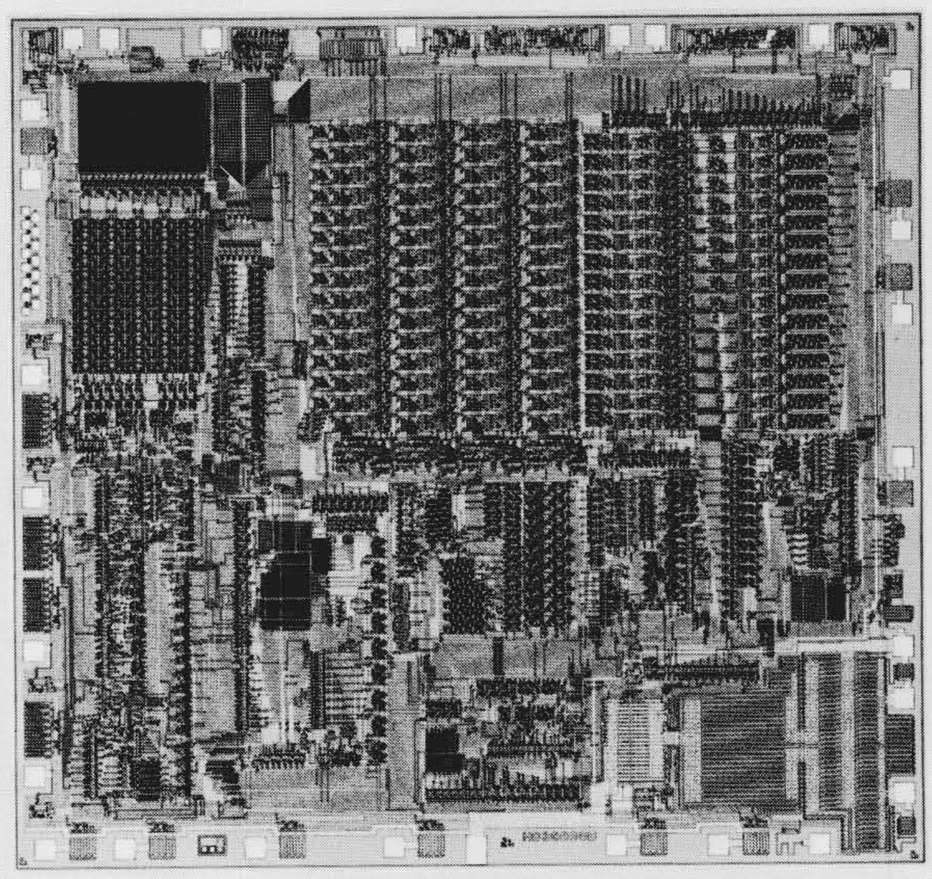

図5　音声合成LSI　15ビット×10ビットの乗算器，加減算器，ディジタルアナログ変換器などから成り約5,000ゲート，チップサイズは約6mm×5.5mm，プロセスはPMOSである。

因である鼻音などの合成時に生ずる振幅異常に対し，声道演算に損失項を加えることにより音質悪化を防止している。また，声道フィルタの演算段数を男性10段，女性8段と切り換えた。フィルタの段数は声道の長さと考えられ，女性は男性に比べ短いので8段とし，その分だけ各フィルタに与えるパラメータのビット数を増やして，音質の向上を図っている。

(2) 機能向上

産業用から民生用までの広い応用分野を考慮して，音声のデータ量は2.4k，4.8k，9.6kビット/秒と自由に設定できる構成とし，顧客の目的とする合成音の量，品質に対応できるように配慮している。また，特殊な機能として，音の高さを変えずに発声速度だけを−60〜＋30%の10ステップに変えられる機能を付加した。これは，教育機器などには効果的な機能である。

4 応用分野

音声合成は，通信用，コンピュータ用，産業機器用，自動車用，教育機器用，時計用，家庭電気品用，玩具用と多くの分野で利用が見込まれている。その中でも応用が多いと思われるものを選んで，**図6**に示す。以下，詳細に述べるエレベーター・エスカレーター用への応用のほかに，自動車用の警報器，コンピュータ端末などへの利用は効果も大きく，実用化は近いと考える。

5 エレベーター・エスカレーターへの応用例

上述した音声合成LSIによる音声合成装置の応用として，エレベーターへの実施例を中心に述べる。

不特定多数の乗客にサービスするエレベーターは，その性能・機能は年々高まってきているが，マンマシン性の一段の向上を図るため，自動放送装置が注目されていた。

主なねらいは，エレベーターの位置及び運転状態を検出した上で行なう通常運転のサービス案内のほかに，地震，火災

図7 音声合成自動放送装置を設けたエレベーター　都内のビルに納入したもので，エレベーターの運転案内放送などにより利用者から好評を得ている。

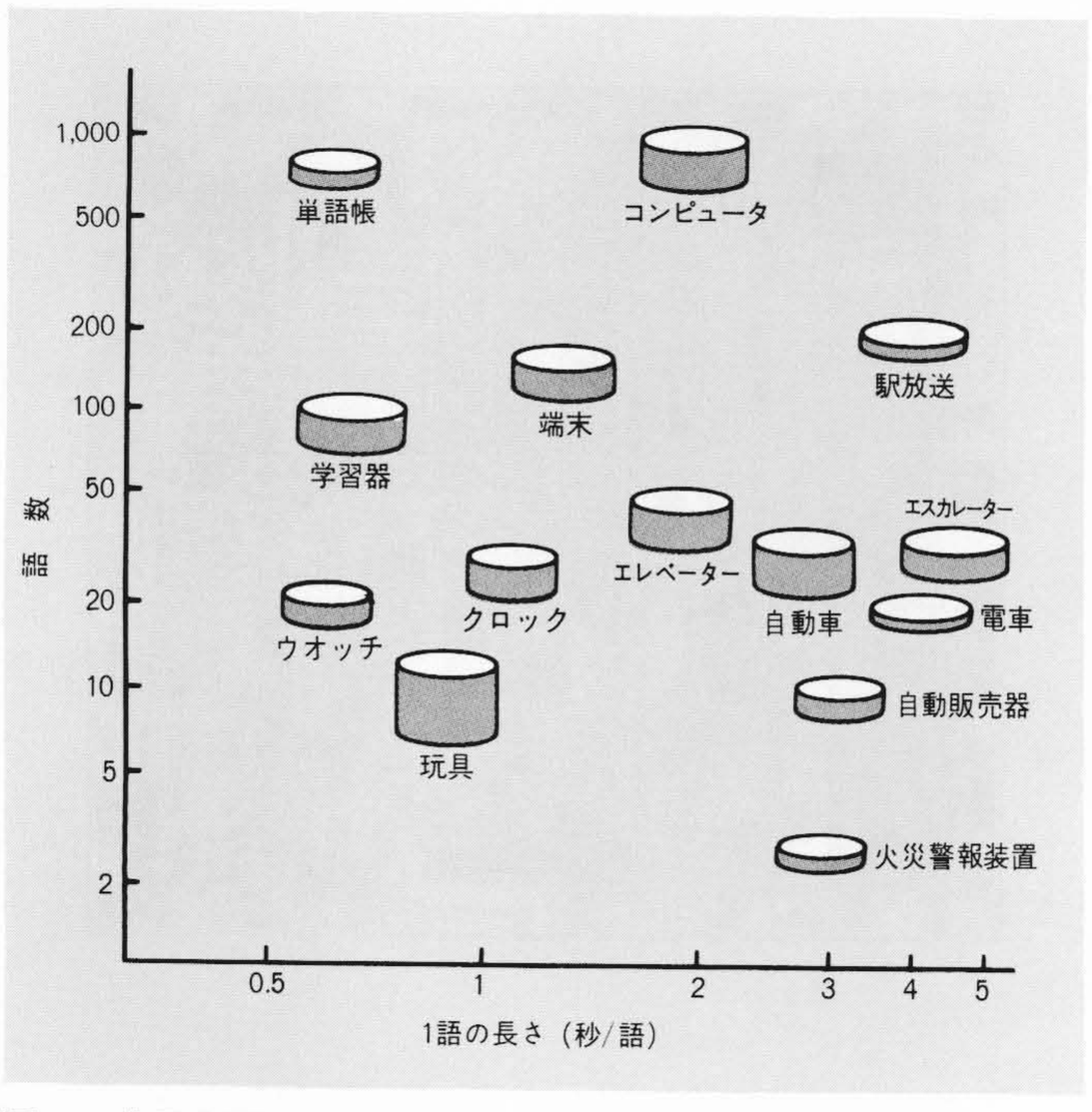

図6 応用分野と音声の規模　自動車用の警報器，コンピュータ端末，エレベーターやエスカレーターなどは効果が大きい。

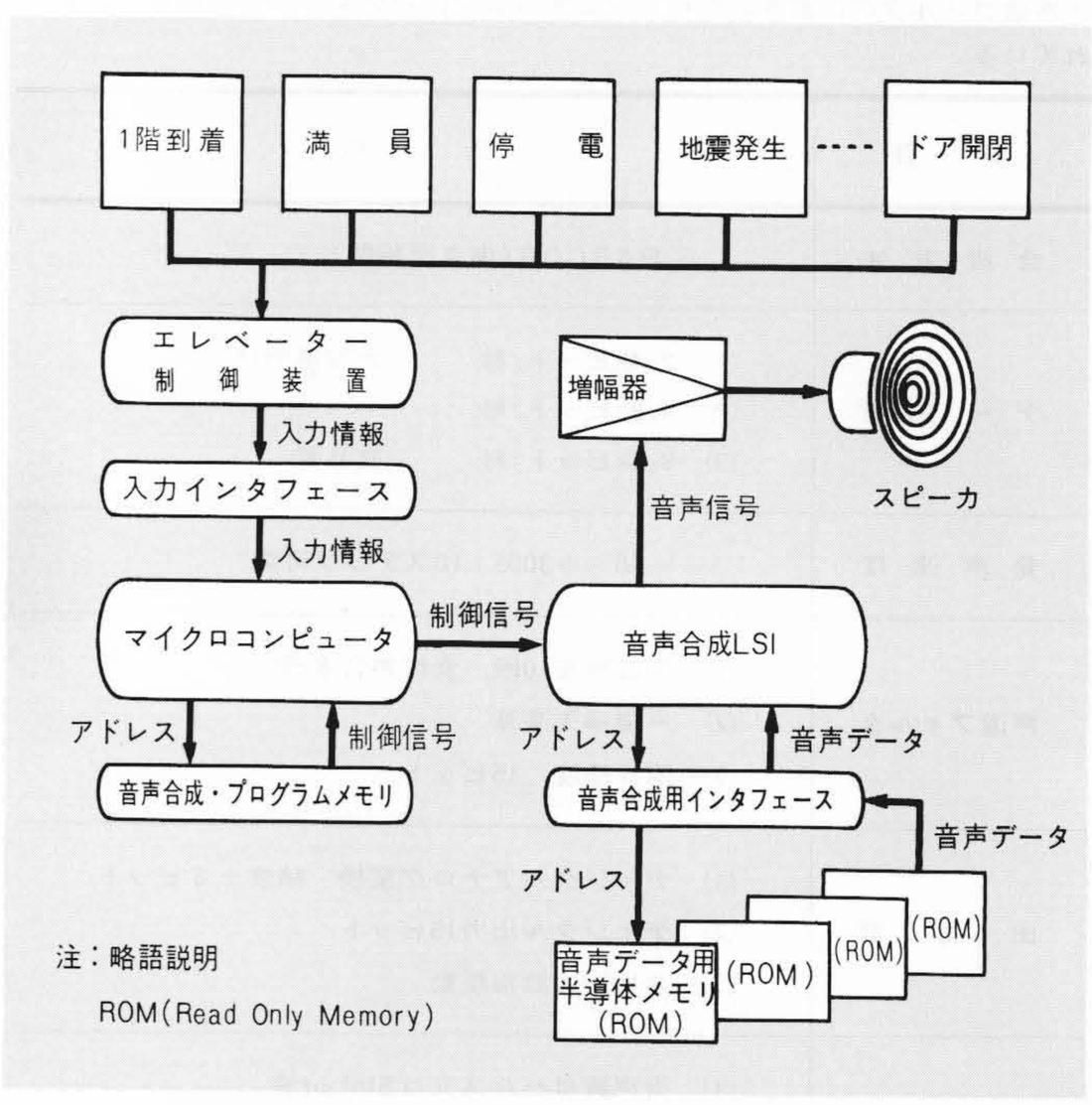

図8 音声合成自動放送装置構成図　エレベーター用として実用化した音声合成自動放送装置を示した。マイクロコンピュータで音声合成LSIが制御される。音声データは半導体メモリ(ROM)にすべて収録されている。

時などで適切なセンサとの連動を図り，管制運転へ自動移行あるいはその案内，更に万一エレベーターの故障の場合にもインターホンの取扱いなど，適切な処置法を乗客に音声により指示・案内し，安全性，操作性などの向上を付加するものである。

図7に，都内のビルに納入した自動放送装置を設置したエレベーターを示す。

5.1　装置の構成

音声合成自動放送装置は，**図8**に示すとおり音声合成LSI，マイクロコンピュータ，音声データ用半導体メモリ（ROM：Read Only Memory），各種のインタフェースなどから構成される。

あらかじめ，放送すべき音声（原音）を分析し，抽出した特徴パラメータを，音声データとしてROMに記憶させておく。放送時は，エレベーター制御装置から入力インタフェースを介して，マイクロコンピュータに案内放送の種別と放送するタイミングに関するデータが送られ，そのデータに基づいて選択された放送内容を，音声合成LSIはROMの音声データをもとに音声を合成し，増幅器，スピーカを通して案内放送を行なう。

5.2　処理手順と放送内容

本装置は，前述したようにマイクロコンピュータによって制御される。その主な処理手順は，

(1)　入力信号を読み込む。
(2)　音声合成LSIに初期条件を設定する。
(3)　音声合成LSIに発声指示を与える。
(4)　音声合成LSIの動作状態を監視する。

であり，その全体フローチャートを**図9**に示す。以下，上記(3)の発声指示を与える部分について更に説明する。

音声合成のための情報は音声データ用ＲＯＭに記憶されるが，案内放送の語句（文節）の組合せ情報をROMから読み出すには語句の組合せ情報のアドレスを指定することで可能である。

エレベーター用案内放送に使用する放送文は，あらかじめ幾つかの語句（文節）に区切り，それぞれにインデックスコードを与える。このコードからデータのアドレスを知るために，各コードの語句の組合せ情報がどこのアドレスから始まるかを，**図10**に示すように一覧表の形で記憶させておく。

放送したい語句のコードが入力情報から判別したならば，上述のアドレス表を参照し，その組合せ情報のスタートアドレスを知り，これをもとに音声合成LSIはROM内のデータを解読して目的の放送内容を発声する。

表2に，エレベーター用自動放送装置の標準放送内容と，各々の放送を行なうタイミングを示す。このほかに同表に示していない到着階など，オプションとして全部で24種類（約53秒）の放送が可能である。

5.3　特長と主な仕様

本装置は，音声データ収録も含めてすべて半導体で構成されているため，長期繰返しによる音質劣化のないことはもちろん，機械駆動部がないため，従来のエンドレステープなどを使用した方式に比べて，大幅な長寿命，メンテナンスフリー，小形・軽量化などを実現した。このほか，放送文の頭出しを瞬時に行なう高速アクセスができる大きな特長をもっており，多種類の放送が単一の装置で可能である。

表3に本装置の主な仕様を示す。

5.4　エスカレーターへの応用

デパート，スーパーマーケットなどに設置されているエスカレーターには，正しい乗り方など安全に関する注意放送が広く行なわれている。従来この放送装置は，一般的にエンドレステープ方式が用いられており1日中放送が続けられていることなどから，その寿命，保全性に問題がないわけではなかった。

今回実用化した音声合成自動放送装置は，これらの問題をすべて解決できるもので，エスカレーターへの応用は，最適

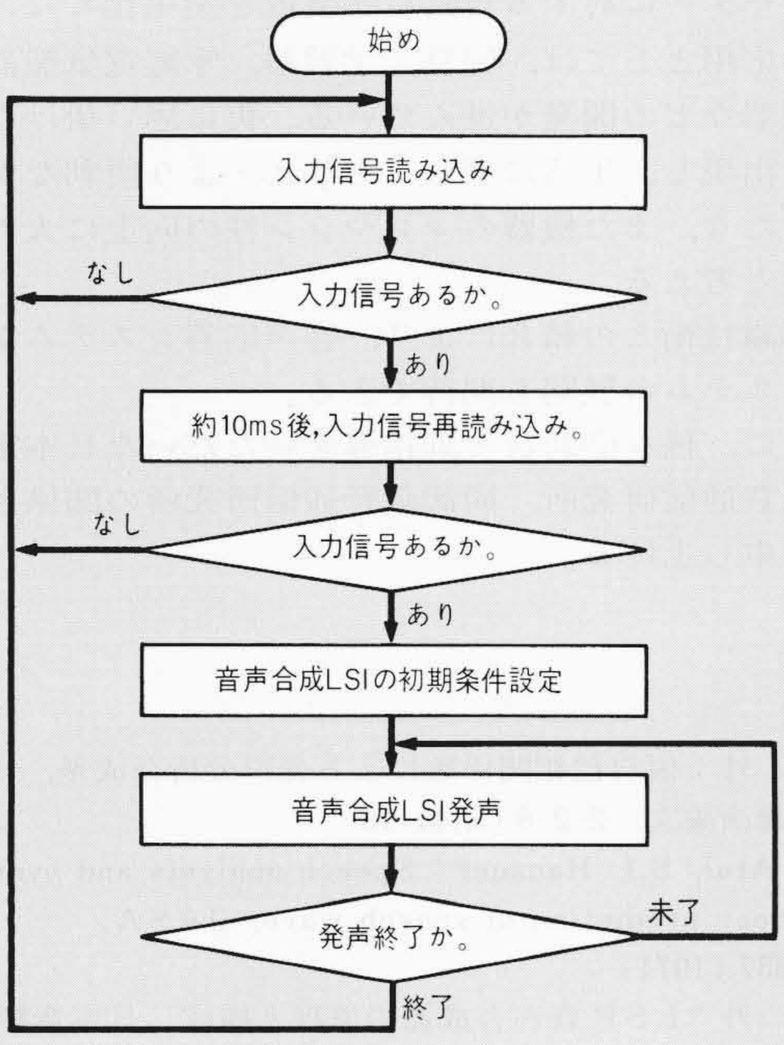

図9　音声合成自動放送装置全体フローチャート　自動放送装置を制御するマイクロコンピュータの処理手順を，マクロのフローチャートで示す。

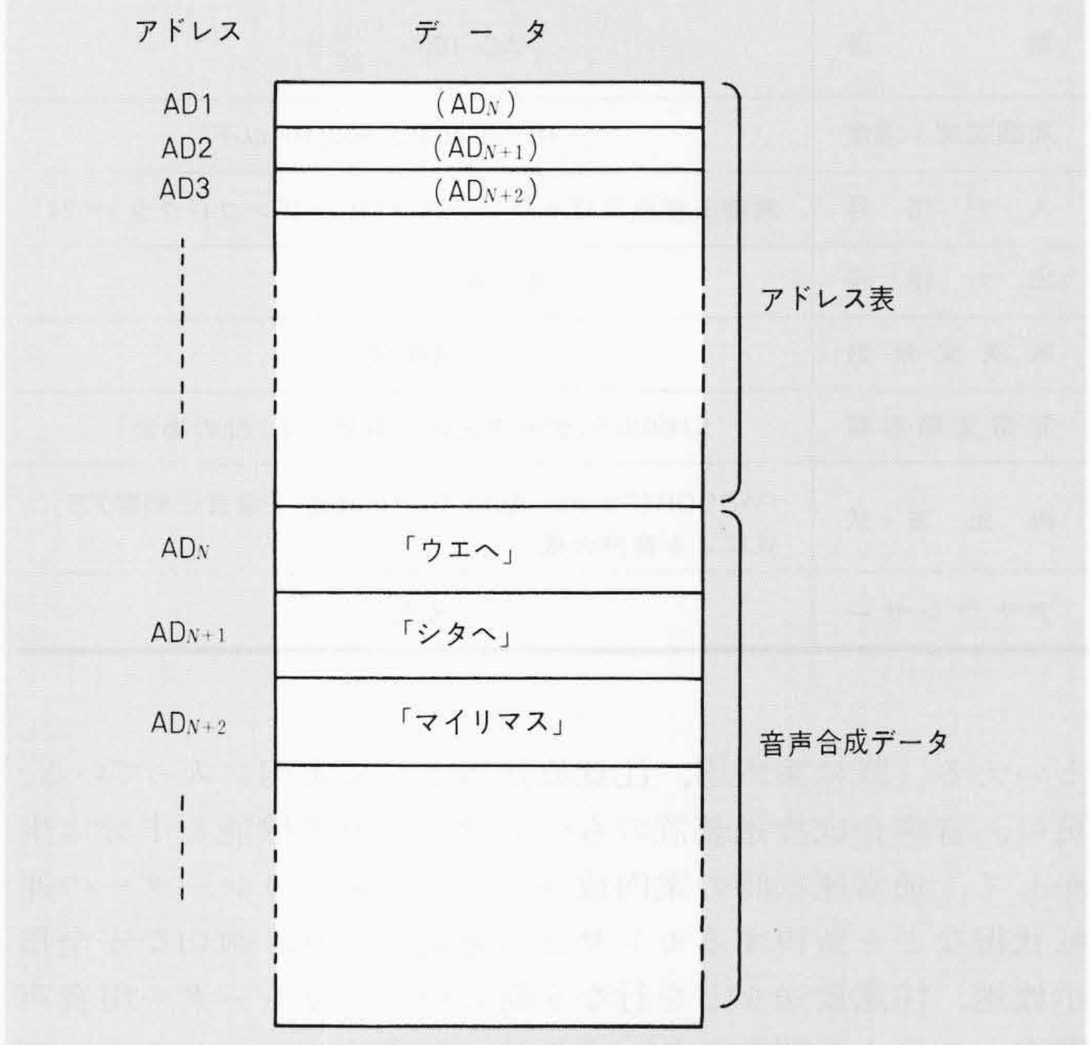

図10　音声データ用ROMの構成　音声データは幾つかの語句に分割し，インデックスコードを付けて，特定のアドレスに収録する。

表2 音声合成自動放送装置標準放送内容 標準放送内容と，その時間，目的及び放送するタイミングを示す。

仕様	No.	放送内容	放送時間	分類	放送タイミング
標準仕様	1	上へ参ります。	約2秒	ホール行先方向案内	かご到着開扉完了後，放送。 ホール呼びリオープン開扉後，放送。
	2	下へ参ります。			
	3	地下2階でございます。	約3秒	かご内到着階案内	かご到着前に放送し，放送終了後 1～3秒で戸開きを開始する。
	4	地下1階でございます。			
	5	1階でございます。	約2秒		
	6	2階でございます。			
	7	3階でございます。			
	8	4階でございます。			
	9	5階でございます。			
	10	6階でございます。			
	11	7階でございます。			
	12	8階でございます。			
	13	9階でございます。			
	14	満員です。後からお乗りの方はお降りください。	約5秒	ホール乗込注意案内	110%過負荷検出で，放送。
	15	ドアが開いたらエレベーターから降りてください。	約5秒	かご内管制案内	火災，地震，自家発電管制運転始動後，放送。
	16	インターホンの呼びボタンを押してください。	約4秒	かご内異常案内	かごが，ドアゾーン以外で停止し，戸が開かないとき放送。
	17	ドアが閉まります。	約2秒	戸閉め注意放送	ドアタイムカウント後，放送終了して戸閉めとする。

表3 主な仕様 音声合成自動放送装置の主な仕様を示す。入力信号として，無電圧接点又はトランジスタ(オープンコレクタ)でインタフェースできるため，汎用性がある。

項目	仕様
音声合成LSI	HD38880
メモリ	PROM HN462716
電源	AC 100V $^{+10}_{-20}$%
周囲温度・湿度	−10～+40℃，90%RH以下
入力信号	無電圧接点又はトランジスタ(オープンコレクタ)×24
出力信号	8Ω負荷 1W以上
放送文章数	24種類
記憶文節容量	53秒以下(データ圧縮2.4kビット/秒の場合)
再生方式	PARCOR(Partial Auto Correlation：偏自己相関)方式による音声合成
アナウンサー	女性

といえる。既に案内用，注意放送用として実用に入っている。更に，音声合成放送装置のもつ高速アクセス機能を十分に生かして，通常運転時の案内放送のほか，エスカレーターの運転状況などを監視するセンサとの連動により，適切な安全指示放送，注意放送などを行なう新しいエスカレーター用音声案内システムを開発した。これは，従来のエンドレステープ方式と比較して，内容的に大きく飛躍したマンマシン性に富んだエスカレーターシステムを可能にしたもので，今後の需要増大が期待できる。

6 結言

数年前まで音声合成技術は，民生，産業分野には無縁のものと思われていたが，LSI技術と結び付いて実用化の気運が一気に高まってきた。日立製作所の最初の応用製品としての，珠算の読上算練習器に続いて，本稿で紹介したエレベーター，エスカレーターに対する自動放送装置を実用化した。

今後の応用としては，玩具，学習器，家庭電気製品，自動車用警報器などの開発が進んでいる。更に思い掛けない応用が次々と出現し，生活に楽しみを与え，より便利なサービスを提供したり，また機器のマンマシン性の向上に大きく貢献するものと考える。

音声認識技術との結合により，音声応答システムなど更に新しいシステムの展開も期待できる。

終わりに，種々御助言と御指導をいただいた日本電信電話公社横須賀通信研究所，同武蔵野通信研究所の関係各位に対して深謝申し上げる。

参考文献

1) 板倉，外：偏自己相関係数による音声分析合成系，日本音響学会講演論文，2-2-6(昭44-10)
2) B.S. Atal, S.L. Hanauer：Speech analysis and synthesis by linear prediction of speech wave, JASA, 50, 637 (1971)
3) 板倉，外：LSP音声合成器の原理と構成，日本音響学会音声研究会資料，S79-46(昭54-11)
4) 嵯峨山，外：複合正弦波による簡易な音声合成法，日本音響学会講演論文，3-2-3(昭54-10)