

音声合成LSI

Speech Synthesis LSI

日立製作所では、昭和54年にPARCOR方式によるPMOS音声合成LSIを開発し、以後各種分野への応用を進める一方その改善を行なってきた。このたび、簡易システム用、低消費電力の1チップCMOS音声合成LSI(HD61885)を開発した。このLSIは、3 μ Hi-CMOS技術により設計され、32kビットのROM、キーボードインタフェース内蔵により、スタンドアロン動作が可能となった。更に、標準化周波数を高くし、積分音源の使用により高音質の合成音を実現できた。

本稿では、1チップCMOS音声合成LSIを主体に、日立製作所の音声合成LSIについて紹介する。

鳥羽喜富* Yoshitomi Toba

鳥居周一** Shûichi Torii

原 英夫** Hideo Hara

1 緒 言

1978年(昭和53年)米国T I (Texas Instruments) 社から音声合成LSI (TMC-0280)及びその応用製品である話す学習器“Speak & Spell”が発表された¹⁾。これは音声合成技術を身近なものとし、各種分野へ適用可能となることを示唆した点で大きな話題を提供した。以後、国内外で音声合成LSIの発表が相次いでいる。日立製作所では昭和54年に日本電信電話公社の指導により、国内で最初に、PMOS (P Channel Metal Oxide Semiconductor) プロセスによる音声合成LSI (HD38880)の開発に成功し、以後、珠算の読上算練習器、音声時計付ラジオ、エレベーターの音声案内²⁾、自動車警報装置、自動販売機、音声端末装置など各種の分野に応用してきた。このたび、より広範な応用分野を目指し、かつその後の音質改善の成果を加えた1チップCMOS (Complementary Metal Oxide Semiconductor)[※] 音声合成LSI (HD61885)の開発を行なった。

本稿では、まず音声合成の原理を述べ続いて日立製作所の音声合成LSIについて紹介する。特に、1チップCMOS音声合成LSIでの音質の改良点について紹介する。

2 音声合成の各種方式³⁾

図1に、現在実用化あるいは考案されている音声合成の方式を、その必要データ量との対比で示す。音声合成方式は3種に大別できる。音声の信号波形をそのまま符号化する波形符号化は、現在電話回線で使われているPCM(パルス符号変調)、非線形PCM、ADPCM(適応差分PCM)などがあり、良好な音質を得ることができる反面、膨大なデータ量を必要としている。一方、原音声からその特徴パラメータを抽出して、それにより合成する分析合成方式は、少ないデータ量で比較的良好な音質が得られるために、現在の音声合成LSIの主流となっている。分析合成方式のうち、LPC(Linear Predictive Coding: 線形予測符号)の一種であるPARCOR(Partial Auto-Correlation: 偏自己相関係数)方式は、その特徴パラメータの抽出が比較的容易であり、高音質が得られる点から広く使用されている。ホルマント合成は、その特徴パラメ

ータとして原音声のホルマント周波数を用いるものであるが、その抽出が困難なためにあまり使われていない。データ量を更に圧縮する方式として規則合成がある。この方式は、原音なしで規則により音声合成するものであるが、現状では良好な音質を得ることが困難である。

3 PARCOR方式^{4),5)}

分析合成方式のうちLPC方式は、音声信号から自己相関係数を抽出してデータ量を圧縮するもので、古くから研究されていた。PARCOR方式はLPC方式の一種であるが、特徴パラメータとして偏自己相関係数を用いるものでLPCよりも系が安定なこと、パラメータが互いに独立で扱いやすい点から現在広く使われている。

PARCOR方式を含め一般に音声合成の回路モデルは、図2に示すように人体の発声機構をシミュレートしている。人体では母音や“b”、“d”のような有声音の発声に際して、声帯の振動により気流を生じ(音源と呼ぶ。)、その気流が咽、口、鼻

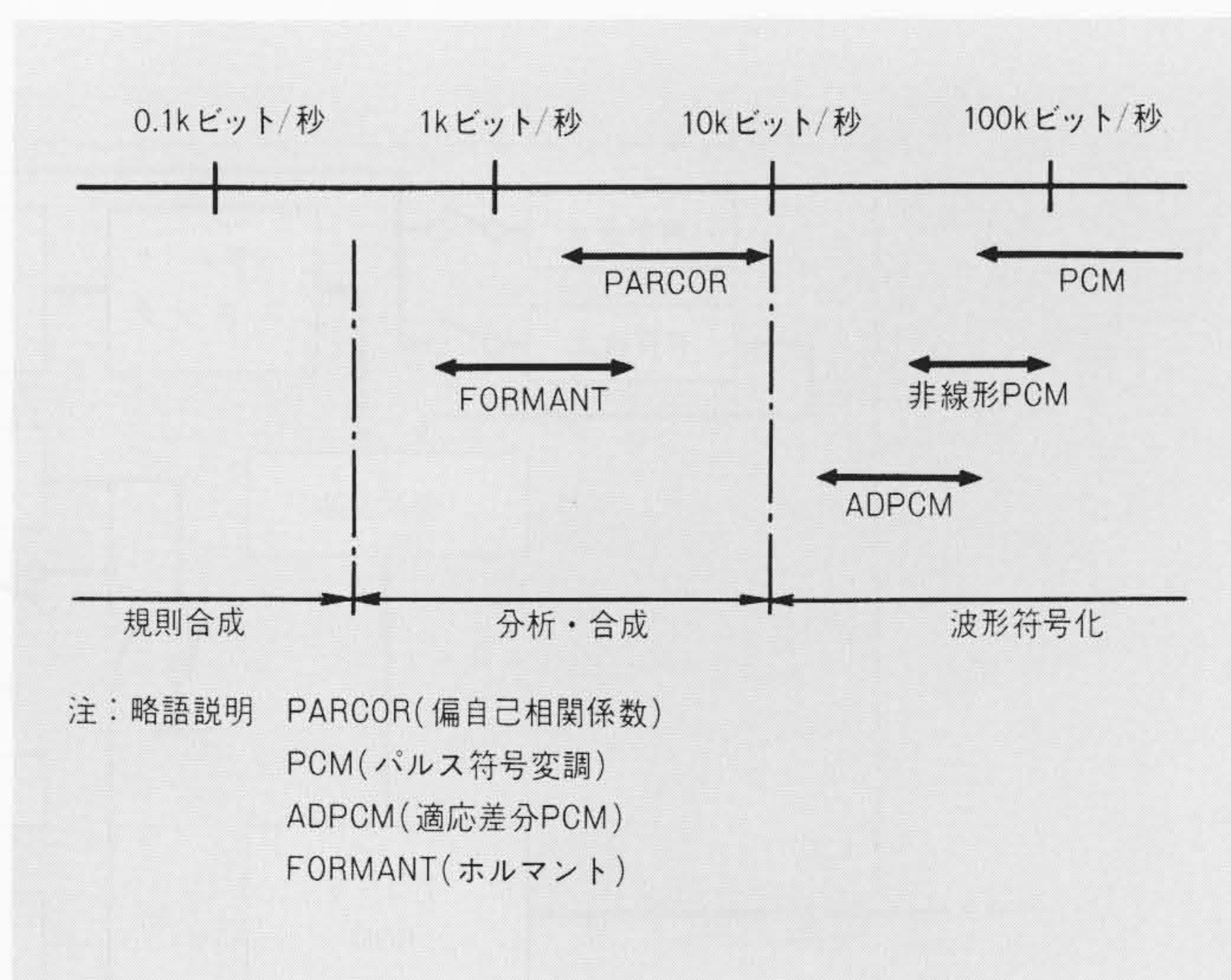


図1 音声合成の各種方式 各種音声合成方式とその必要とするデータ量を示す。

※) CMOSとは半導体製造法の一つで、低消費電力で高集積度が得られる方法である。

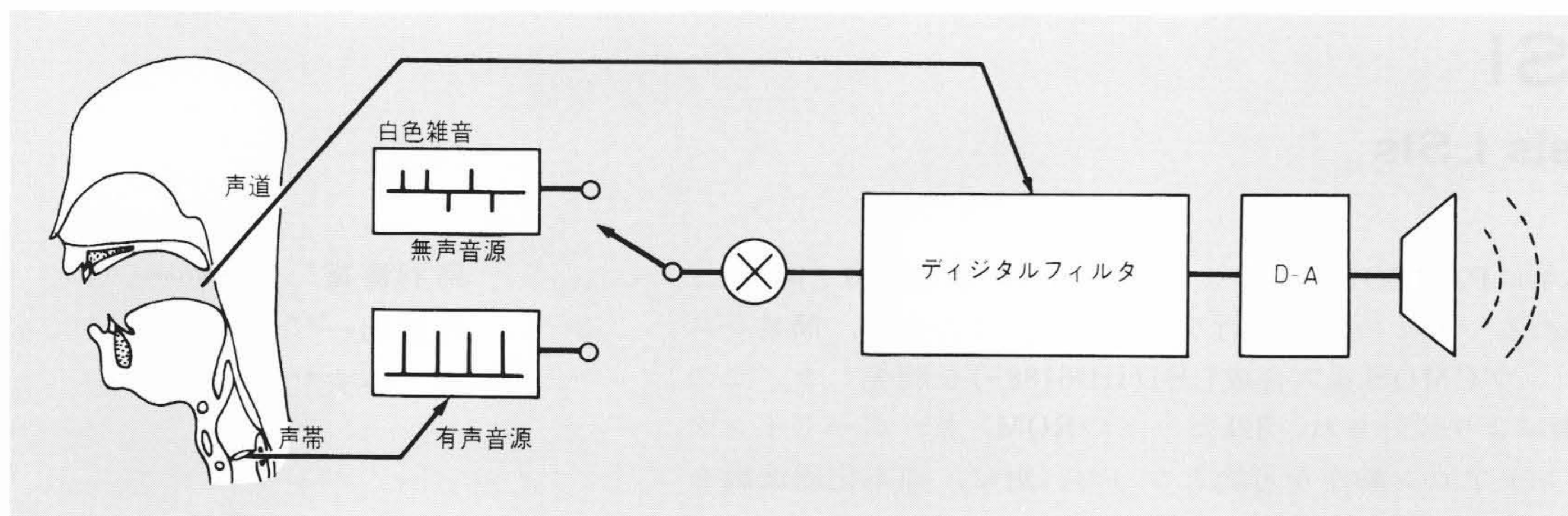


図2 PARCOR発声モデル PARCOR音声合成では、声帯は音源回路、声道はデジタルフィルタでシミュレートされる線形モデルとなっている。

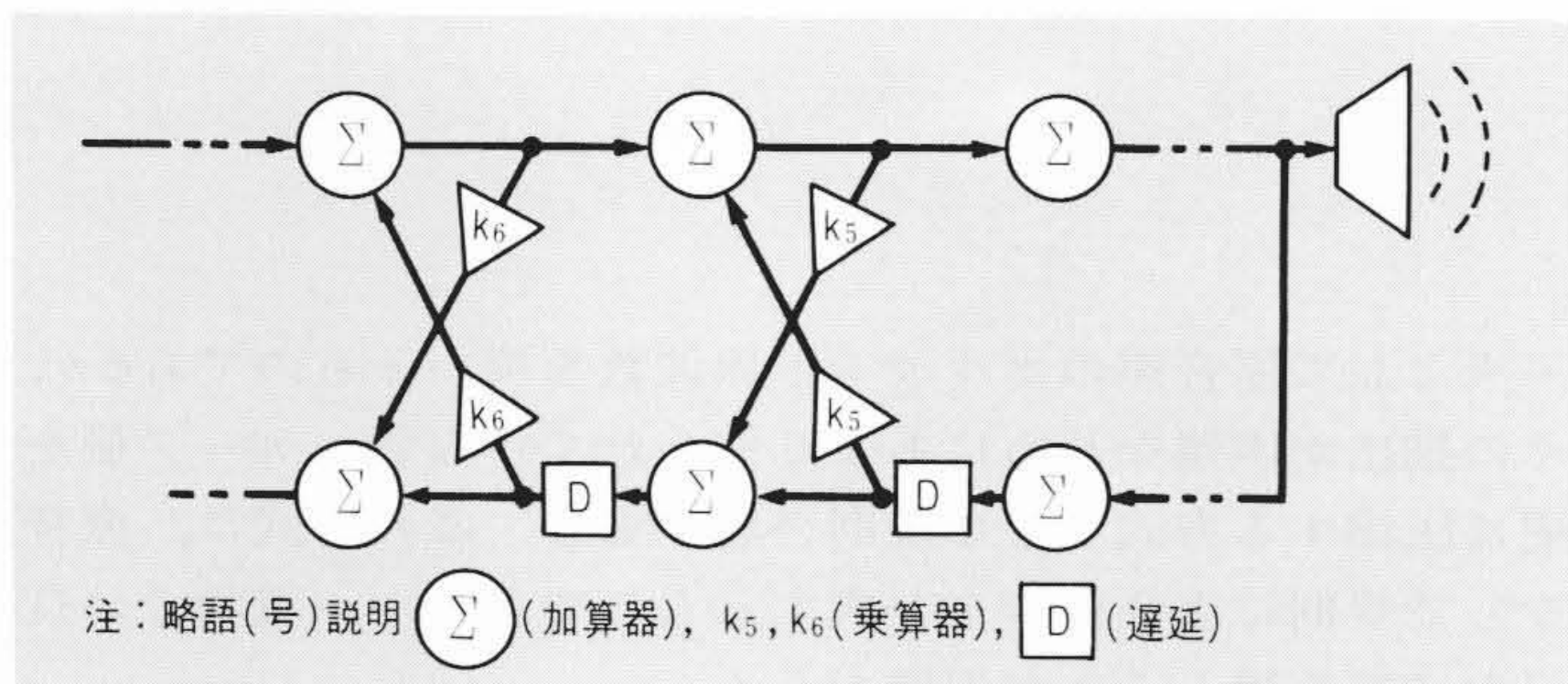


図3 PARCORデジタルフィルタ 格子形フィルタを8段若しくは10段縦続接続により、構成されている。

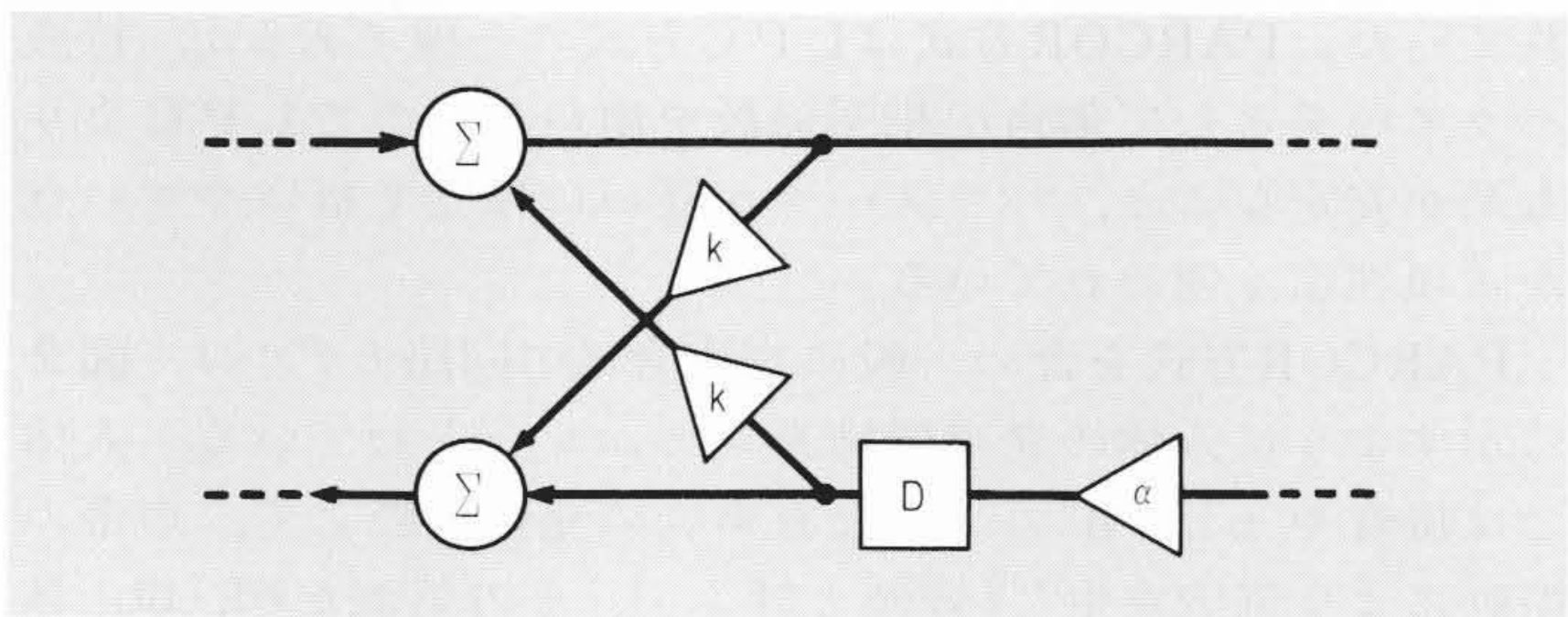


図4 声道損失挿入フィルタ回路 反射波を $\alpha(0.992)$ 倍することにより、声道損失効果を実現する。

(声道と呼ぶ。)を通る間に音韻が形成される。また“p”, “t”のような無声音では、声道内に生じた乱流が音源となる。PARCOR方式では、音源部を有声音では周期的インパルスが発生するプログラマブルカウンタで、無声音では白色雑音で近似している。更に、声道は二管モデルのデジタルフィルタで置き換えている。デジタルフィルタは、図3に示すような2掛算の格子形フィルタの縦続接続により構成されている。ここで各格子形フィルタは、その係数(k パラメータ)で制御される。 k パラメータは、原音声から分析・抽出される。他に音声分析時には、有声音・無声音の判別、有声音時には声帯の振動周波数(Pitch: 音声の高さを示す。)及び音声の強さを示すパラメータ(Amp)が抽出される。ただ女声音ではPitchが高く、ホルマントのバンド幅が過少に推定される。この対策として音声合成時にバンド幅を広げるために、フィルタに損失をもたせる声道損失効果を挿入すると有効である。日立製作所の音声合成LSIでは、デジタルフィルタは図4に示す損失を含めた格子形フィルタが基本となっている。

4 音声合成LSI

PARCOR音声合成システムでは、デジタルフィルタの演算のため、1標本値間($125\mu s$)に20回の乗算と加算が必要である。これをLSI上で実現するために、1個のパイプライン乗算器と加算器及び遅延素子を時分割的に切り替えて使用

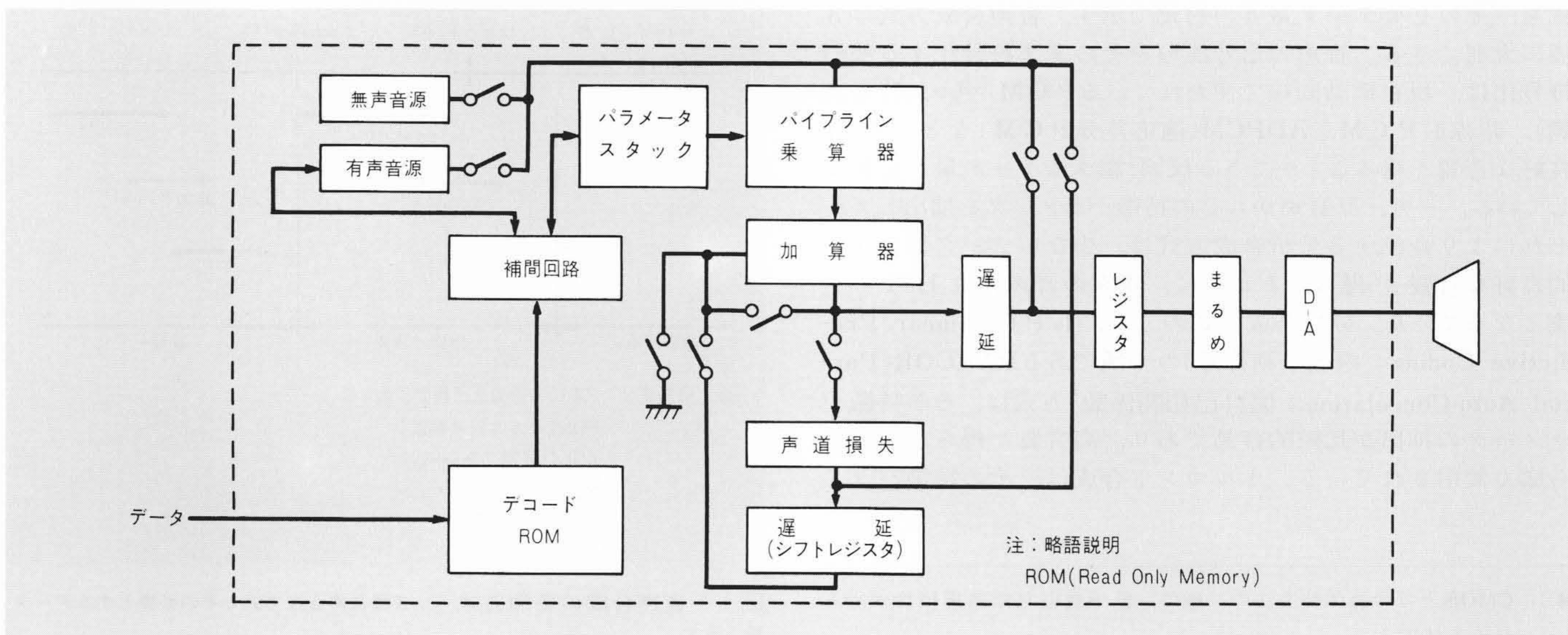


図5 音声合成LSIブロック図 デジタルフィルタは、1個乗算器と加算器を遅延素子と組み合わせて、時分割に使用して実現している。他に音源回路、パラメータの補間回路などから成る。

する。一方、音源回路として有声音用のプログラムカウンタ、無声音用の白色雑音発生器がある。これらを制御するパラメータ(Amp, Pitch, $k_1 \sim k_{10}$)は、音声データとし10~20msごとに更新される。このパラメータの更新時の急激な変化によるひずみの発生を防ぐために、パラメータの変化を滑らかにする線形補間回路がある。他に音声データ(各パラメータ)は、その分布により非線形に符号化されているために、それを戻すデコードROM(Read Only Memory)、最終的に音声信号を得るD/A変換器などがある。図5に音声合成LSIの基本ブロック図を示す。

表1 日立音声合成システム PMOSシステムとCMOSシステムの比較表で、音源、標準化周波でCMOSシステムは音質の改善を行なっている。

項目 \ 区分	PMOS	CMOS																														
システム構成																																
ROM 容 量	128kビット/外部ROM	32kビット/内蔵ROM 128kビット/外部ROM																														
ビットレート	<table><tr><td colspan="2"></td><td colspan="2">フレーム長さ</td></tr><tr><td colspan="2"></td><td>20</td><td>10</td></tr><tr><td rowspan="2">ビット / フレーム</td><td>48</td><td>2.4</td><td>4.8</td></tr><tr><td>96</td><td>4.8</td><td>9.6</td></tr></table>			フレーム長さ				20	10	ビット / フレーム	48	2.4	4.8	96	4.8	9.6	<table><tr><td colspan="2"></td><td colspan="2">フレーム長さ</td></tr><tr><td colspan="2"></td><td>20</td><td>10</td></tr><tr><td rowspan="2">ビット / フレーム</td><td>50</td><td>2.5</td><td>5.0</td></tr><tr><td>99</td><td>4.95</td><td>9.9</td></tr></table>			フレーム長さ				20	10	ビット / フレーム	50	2.5	5.0	99	4.95	9.9
		フレーム長さ																														
		20	10																													
ビット / フレーム	48	2.4	4.8																													
	96	4.8	9.6																													
		フレーム長さ																														
		20	10																													
ビット / フレーム	50	2.5	5.0																													
	99	4.95	9.9																													
標準化周波数	8kHz	10kHz																														
音 源	有声音：インパルス 三角波 無声音：M系列符号	有声音：積分音源 無声音：M系列符号																														
ディジタル フ ィ ル タ	8段又は10段 声道損失：α=0.992 乗算器 10ビット×15ビット	10段 乗算器 10ビット×15ビット																														
コード化 則	k ₁ , k ₂ ：tanh ⁻¹ k k ₃ ～k ₁₀ ：線形	k ₁ , k ₂ ：sinh ⁻¹ k k ₃ ～k ₁₀ ：線形																														

注：略語説明
PMOS(P Channel Metal Oxide Semiconductor) MC(マイクロコンピュータ)
CMOS(Complementary Metal Oxide Semiconductor)

表2 日立音声合成LSI 日立製作所の音声合成LSIの一覧表で、CMOSシステムは低消費電力、広い電源動作範囲を実現している。

項目	PMOSシステム			CMOSシステム	
LSI名	HD38880	HD38881	HD38882	HD61885	HD44881
機能	PARCOR形 音声合成LSI	128kビット音声合成 専用ROM	EPROM インタフェース	PARCOR形 音声合成LSI (32kビット音声ROM内蔵)	128kビット音声合成 専用ROM
プロセス	PMOS(5 μ)	PMOS(5 μ)	PMOS(5 μ)	Hi-CMOS(3 μ)	CMOS(5 μ)
電源電圧	-10V	-10V	-10V, -5V	3.6~5V	5V
消費電力	350mW	50mW	400mW	動作時：2.5mW スタンバイ時：5 μ W	動作時：2.5mW スタンバイ時：5 μ W
外形	DILP-28	DILP-28	DILP-42	DILP-28	DILP-28

注：略語説明 Hi-CMOS(High performance CMOS), EPROM(Erasable and electrical Programmable Read Only Memory)

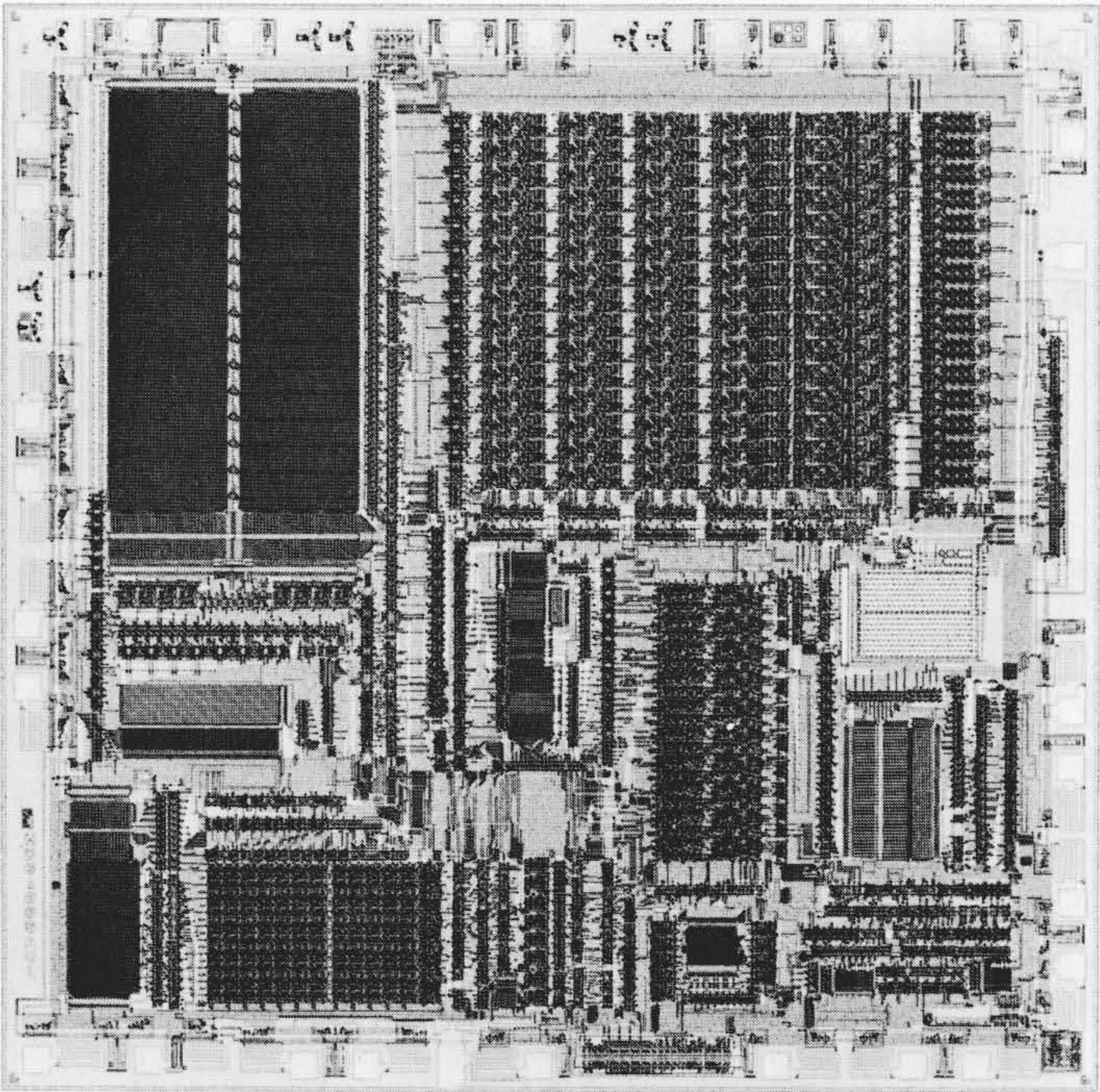


図6 HD61885チップ 1チップCMOS音声合成LSI(HD61885)のチップで、その面積は5.40mm×5.46mmである。

表1に日立製作所の音声合成システムを、表2にそれらを構成するLSIを紹介する。PMOSシステムは、昭和54年に開発されたもので、PARCOR方式音声合成LSIと音声データを記憶するための専用のマスクROMから成る。更に、これら全体を制御するためのマイクロコンピュータが必要である。他に少量生産用及びプロトタイプ用に音声データをEPROM(Erasable and electrical Programmable Read Only Memory)に記憶し、音声合成LSI、EPROMインタフェースから成るシステムも可能である。一方、新しく開発されたCMOSシステムは低消費電力、簡易システムを目標に開発されたもので、チップ上に32kビットの音声データ用のマスクROMを内蔵している。更に、キーボード入力回路も内蔵しているために、マイクロコンピュータなしでの発声も可能である。また、発声語数の増加のためには、専用のマスクROMを容易に付加できる。図6に、この1チップCMOS音声合成LSI(HD61885)のチップを示す。本LSIは3 μ Hi-CMOS(High performance CMOS)技術で設計された高集積度LSIの一種である。

5 音質改善

日立製作所ではPMOS音声合成システムを発表以来、ソフトウェア・ハードウェア両面からその音質改善を検討しており、その結果がCMOS音声合成LSIに反映されている。

5.1 標本化周波数

表1に示すように、標本化周波数は10kHzに上げている。これは図7に示すように、分析時の前置のLPF(低域通過フィルタ)及び合成時のLPFにより、音声スペクトルの高域部が損なわれることによる音質劣化の防止を目的としている。第3ホルマントから4kHzにかけての周波数成分が、音質に

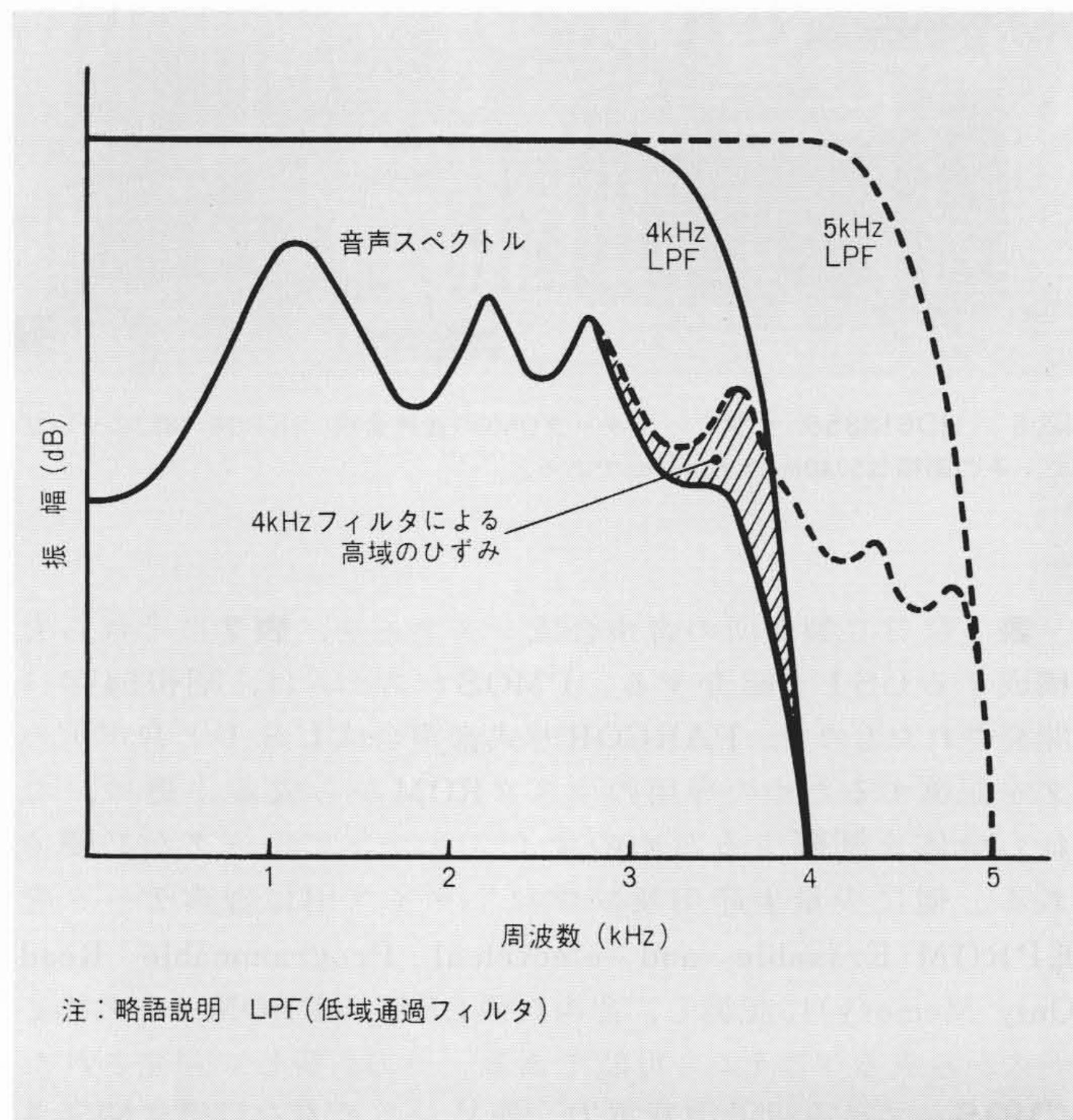


図7 フィルタによる音声スペクトルひずみ 8kHz標本周波数(4kHz LPF)では、高域(第3ホルマント以上)での減衰により、スペクトルにひずみがみられる。

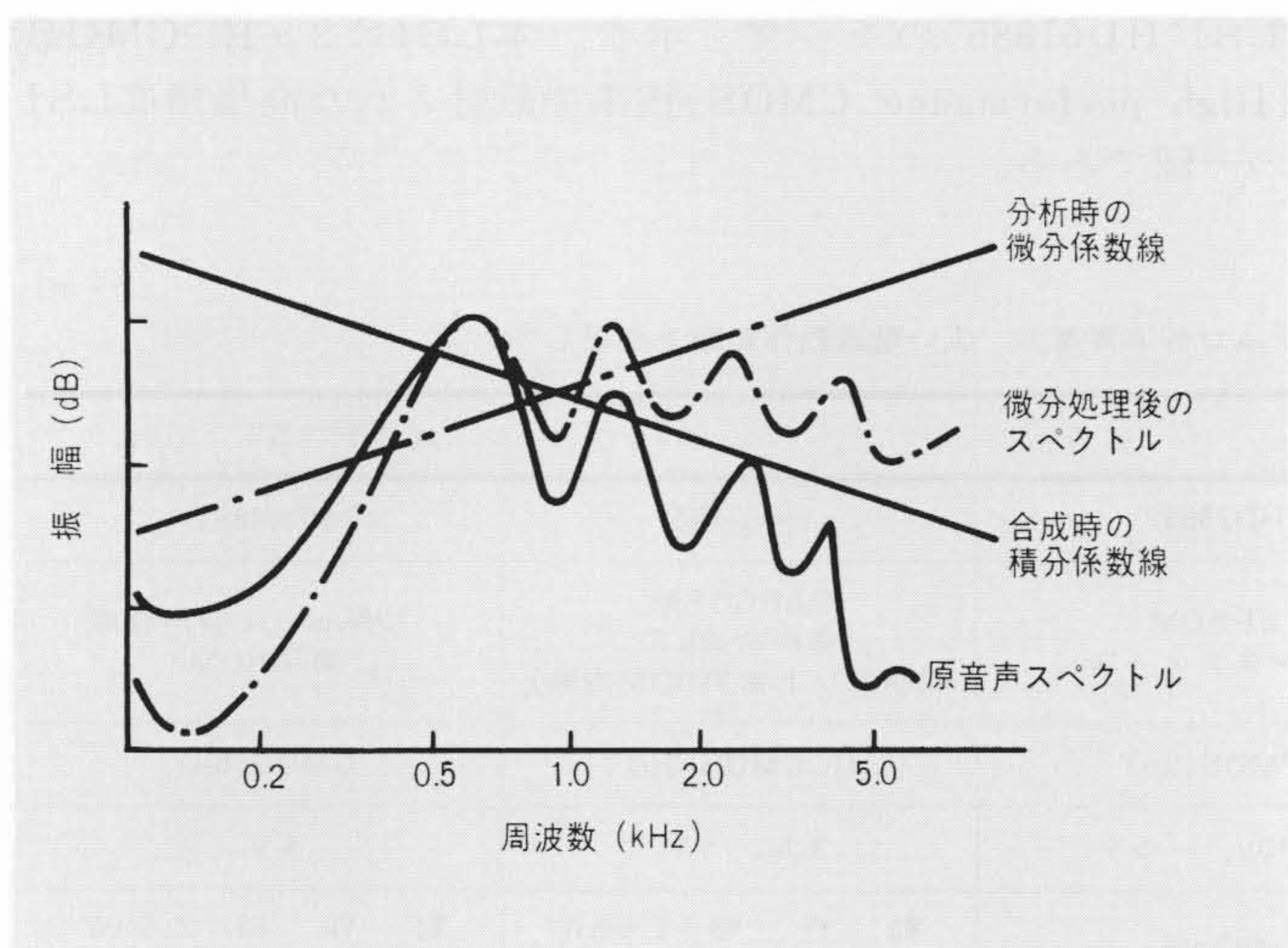
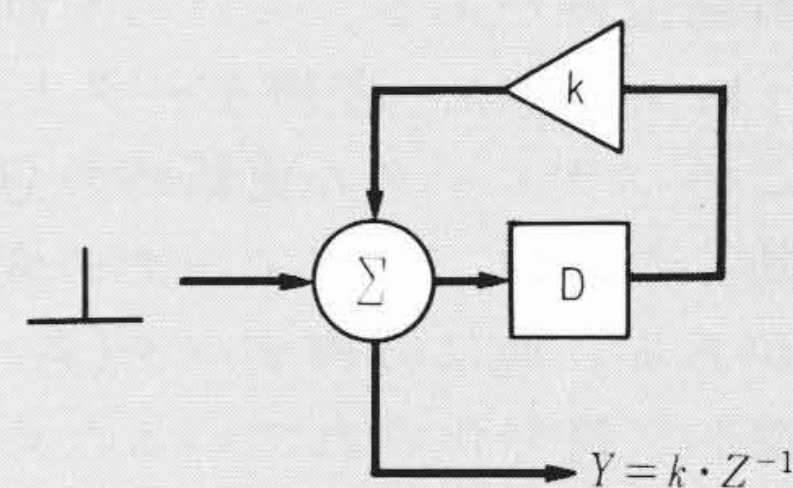
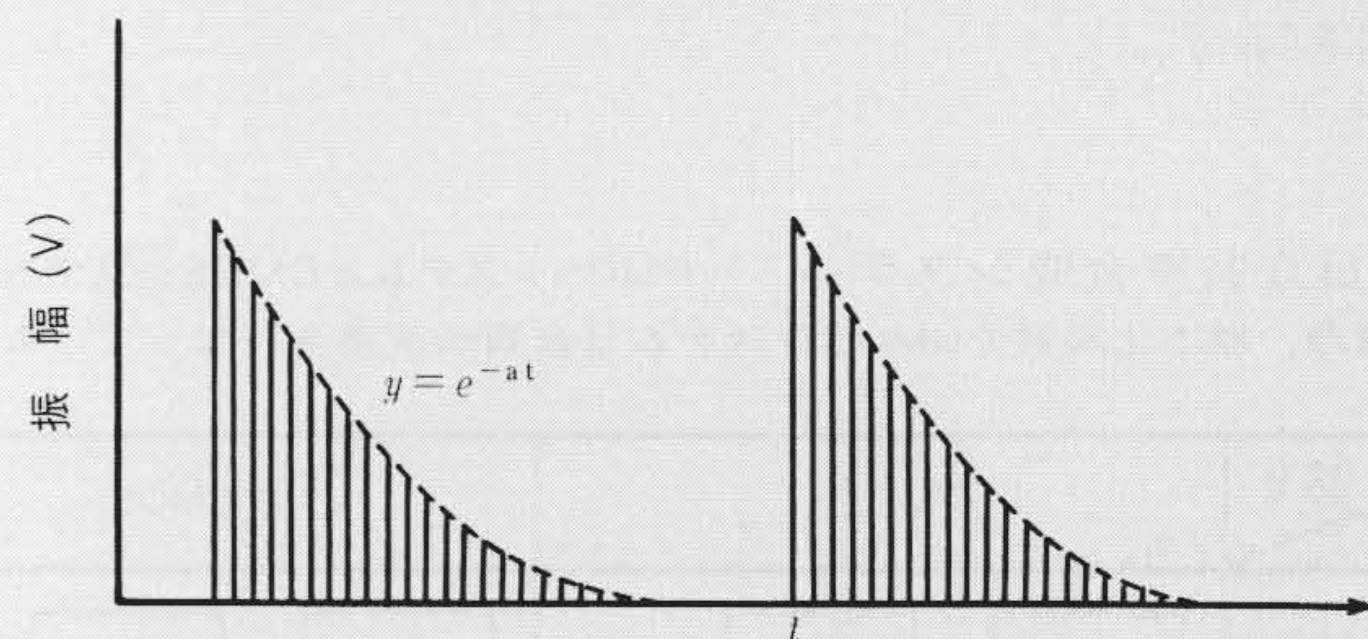


図8 微分・積分によるスペクトルの変化 分析前の微分処理により、高域を持ち上げ分析感度を良くする。更に、合成時には積分音源を使用することにより、再生して全体として自然な音質を得る。



(a) 積分音源



(b) 時間

図9 積分音源と積分音源波形 積分音源は、インパルスを積分することにより得られる。その波形は、指数関数的に減衰する。

かなり影響を与えていることが判明したため、標本化周波数を10kHzまで上げ、4kHzまでの減衰を防いだ。

5.2 積分音源

通常の音声スペクトルは、1kHz以下にエネルギーが集中しているため、ホルマント周波数を推定する音声分析に誤差要因となった。これを克服するために、分析時に原音声信号を微分する方式がとられる。しかし、これにより分析した音声は合成時に積分する必要がある。この関係を図8に示す。しかし、合成音を積分する場合は、外付け積分器が必要となるため、HD61885では音源に積分特性をもたせた積分音源を使用することにより、システムの簡素化を図った。この積分音源は、図9(a)に示すようにPARCORの基本音源モデルであるインパルスを積分して得られ、その波形は同図(b)に示すようになる。

6 結 言

PMOS音声合成LSIの開発により、各種分野へ音声合成LSIの採用を促進した。その後の音質改善により、低消費電力、簡易システムの1チップCMOS音声合成LSIを開発した。これにより時計などの小形機器への音声合成LSIの応用が促進されるものと思われる。

参考文献

- 1) R. Wiggins, et al.: Three-Chip System Synthesizes Human Speech, Electronics, pp. 109~115 Aug. 1978
- 2) 弓仲, 外: 音声合成技術の現状とその応用, 日立評論, 63, 6, 405~410 (昭56-6)
- 3) G. Heftman: Speech Synthesizers—the talk of the IC world, Electronic Design, pp. 73~76, March 5, 1981
- 4) 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)
- 5) 中田和男: 音声, 日本音響学会編 音響工学講座, コロナ社 (昭52-12)