音　声　の　合　成*

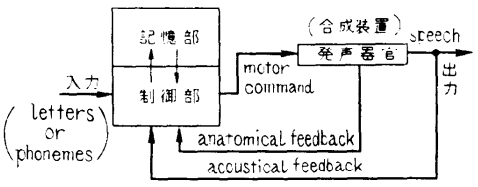
中　田　和　男**

1. はしがき

人間の発声によらないで，機械的に人間の音声を作るいわゆる音声合成の研究には二つの目的が考えられる．その一つは，音声そのものの研究あるいは音声の機械による自動識別の研究のための有力な研究手段としての合成の研究であり，他の一つは，各種の情報処理機械と人間の間の自然で能率のよい通信手段開発のための合成の研究である．音声の識別と合成の間の関係についてはあとにふれることにするが，後者の例としては，たとえば電話ダイヤリングによる問い合わせに，製品在庫の現状を答える計算機とか，各種の問い合わせ業務において再生される情報を，音声の形で答え返す情報検索機械とかが考えられる．

この二つの目的に対して，音声合成の研究のしかたが，本質的には同じでありながら，やや異ってくる．前者では常に人間の発声のできるだけ厳密なシミュレーションであることが要求され，いかにしたらより厳密なシミュレーションにすることができるかを研究することが，合成の側としての研究目標であるのに対して，後者では，たとえ近似的な解であっても，合成された音声が十分の明瞭性とある程度の自然性をもったものであれば，装置の簡単さとか制御の容易さという点ですぐれていれば，実用的な解となりうるし，ある意味ではそういう解法を求めることが工学的研究の目標であるともいえる．

人間の発声機構を機械による音声合成との対応に重点をおいて考えると**第1図**のようになる．このような



第1図　人間の発声機構の工学的な説明

機構を機械的に実現するにあたって記憶部に重点をおき，制御の過程を compilation（編集）のみにしたのがすでに録音されている音声断片からの**編集による音声の合成**（Speech Synthesis by Compilation）であり，合成装置とその制御の仕方に重点をおいたのが**法則による音声の合成**（Speech Synthesis by Rules）である．さらにその中で合成装置の機能を人間の発声器官のアナログ的な回路とし，音声波のもっている構造的な情報を合成装置の機能に内蔵させたのが，いわゆる**模擬音声合成**（Analog Speech Synthesis）である．

また合成過程における計算機の用い方を大別すると，計算機自体を合成装置として用いるもの，すなわち Computer Synthesis or Simulationと，計算機を合成装置の制御のみに用いるもの，すなわち Computer Control の二つがある．

以下音声そのものについての記述は必要最小限にとどめ，計算機の利用あるいは計算機への応用という点に重点をおいて音声合成の研究の現状を紹介しよう．

2. 編集による音声の合成

2.1. 録音された音声断片による合成

不連続な入力の系列，たとえば文字とか音韻記号とかによって書かれた文章から連続的な音声(会話音声)を合成する一つの可能性として，あらかじめ適当な単位の音声断片*（以下片素という）を録音しておき，入力に対応してこれらあらかじめ録音されている音声片素の中から対応するものを一つ一つとり出して，適当に時間的に連結すればよいではないか，ということは原理的にはだれでもすぐに考えつくことであろう．事実1953年にすでに C.M. Harris が building block 合成方式として提案している[1,2]．

しかし，このような方法によって不連続入力（文字または記号の系列）から連続的な音声の合成が可能だと考える背後には，連続音声も適当な単位をとれば音韻的のみでなく，自然性のうえでも時間的に分割可能

* Synthesis of Speech, by Kazuo Nakata (Radio Research Laboratories)
** 電波研究所情報処理研究室

* 普通人間の発声によるが必ずしもその必要はなく，あとにのべるような機械によって合成された音声であってもよい．

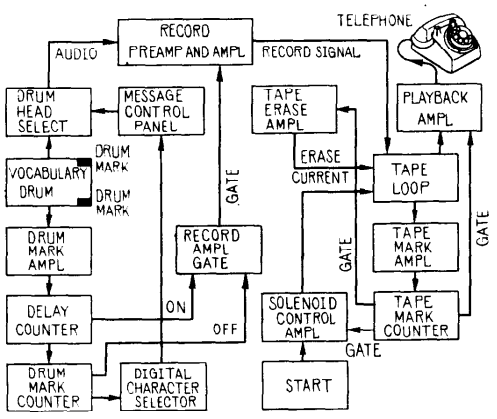
だということを仮定している．この点が実は問題なのであって，人間による音声の聞き取り方という根本的な問題に関連している．

しかし編集による合成が全く不可能だというわけではなくて，実験的に注意深く作られた音声片素の録音素材（inventory という）からの適切な編集によれば，やや不自然に感じられることはあっても，十分了解度の高い連続音声を合成できることが実験的に示されている．ことに合成の語彙（vocabulary）が限られているような場合には，実用上有力な方法といえる．

2.2. 編集による音声合成の実例

（1） IBM の DIVOT (digital-to-voice translator)[3]

この装置のブロック図を**第2図**に示す．この図で



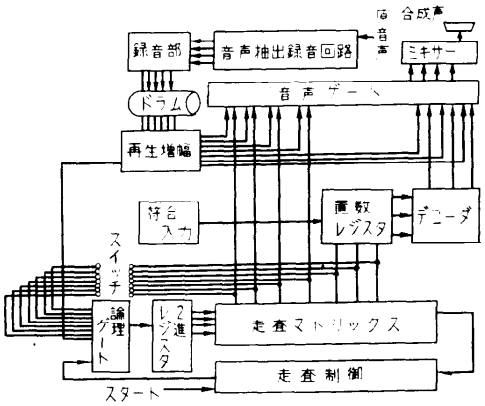
第2図　IBM の DIVOD のブロック図

message control panel に音声として応答すべき message を指定する digital code がセットされ，それに対応した message が高速で vocabulary drum からよみ出され tape loop にうつされる．このような message の assemble（編集）がおわると，この tape が低速に切り換えられて音声として再生される．vocabulary drum は8インチ直径のニッケル・コバルト鍍金のものが用いられ，90 rpm で 300 cps から 3,000 cps の周波数帯域を録音することができる．録音素片の単位の語長は1/3秒か2/3秒で，1/3 秒の短語は1トラックに2回，2/3 秒の長語は二つのトラックに1回ずつ録音されているが，互に 180° づれた位置に録音されており，どの短語と長語の組み合わせも待時間なくよみ出すことができる．よみ出しは 50 倍の高速（4,500 rpm）回転で行なわれるが，そのための品質の低下はみとめられない．

Loop tape は two-speed の buffer として用いられており，録音時には再生時の 50 倍の高速（200 inches/sec）で動作し，たとえば1/3秒の短語15個よりなる5秒間の音声は 100 msec で録音される．したがって loop tape の数さえそろえれば50個の独立情報を同時に音声として再生することができる．実験の結果，綜合特性として S/N 比 30 db，9語よりなる80個の message の30人による聞き取り試験の結果わずかに 0.1%/word 以下の誤りにすぎなかったと報告されている．

（2） 電電公社電気通信研究所の電話番号案内用音声編集装置[4]

これは電話番号の問合わせに対する自動解答装置の一部として研究用に試作されたものであり，そのブロック図を**第3図**に示す．この装置では vocabulary drum の1トラックには1数字語音が録音されてお



第3図　電電公社電気通信研究所の電話番号案内用音声編集装置のブロック図

り，1回転ごとに同期してよみ出されている．慣用句の部分（この例では「ソノ番号ハ」）は固定順序でよみ出されるが，番号数字の部分は置数レジスタとデコーダを介して選択される．この研究用の試作装置で，特に7桁番号（局番3桁で加入者番号4桁）の合成におけるピッチ（アクセント）パタンの選び方と録音音声素片の数をふやすことの効果などについて検討が行なわれた．

2.3. Campiled Speech の問題点

（1） 録音素片の単位とその必要数

編集による音声合成において最も重要な問題は，予じめ録音しておく音声素片の単位のえらび方と必要な

素片の数（inventory）の検討である．一般的にいって素片の単位が大きくなれば，それだけ合成された音声の明瞭度と自然性は向上するが必要な素片数は大きくなり，編集に時間がかかるようになる．英語について Peterson 等が検討した結果の一例を**第1表**に示す[5]．

第1表 録音素片の単位とその必要数

素片の単位	Phoneme Sequence の数	必要素片数
Phoneme	37	155
Phoneme dyad	1,218	8,460
Half-syllable	1,647	11,529
Syllable	4,400	30,800
Syllable dyad	858,458	40,173,336
Word	10,119	

注）Phoneme Sequence の数と必要素片数（inventory size）が違うのは，語中の位置や文法上の働きの違いによる variations を加えて自然性を増しているためである．なお，これらの数値は推定の基礎となる言語統計のとり方によってかなりの違いがある．ここでは Dewey によるものを示す．

この表で dyad を単位とすると phoneme, syllable, word などの音韻的または言語的な単位を用いた場合にくらべて inventory が過大になり，非能率的であるということがわかる．このような検討の結果，おそらく word が最適な単位であろうといわれている．

（2） Spelling の問題

Spelling というのは録音素片の inventory にない入力の処理法として最も基本的な小単位（たとえば phoneme とか syllable とか）の素片からそれを編集合成することである*．Inventory の大きさが有限である以上，編集による合成法で一般的な連続音声を合成しようとすると，この問題を避けることはできない．その確率は，これも英語の場合についてであるが，word を単位とした場合，inventory の大きさが 7,000 語のとき約 5%，20,000 語のとき約 1% と推定される．しかし，この推定では地名，人名などの固有名詞は除かれているので，もしそれらも含めるとすれば実際にはかなり頻度の高い問題と考えなければならぬ．

その他，金物的には大容量で random access の記憶装置とその高速制御ということが必要である．

3. 法則による音声の合成

3.1. 模擬音声合成（Vocal Tract Analog 型）

（1） 声道内音波の波動方程式

人間の発声機構を音響理論的に見れば，音声は音源（声帯の振動（有声音）または空気の乱流（無声音））によって励振された共鳴系（声道（vocal tract））の放射系（唇）からの放射出力と考えられる[7]．したがって音声を物理的にあらわす音波のモードを平面波に限るとすれば**，声道内の音声波の仮播は（1）式の波動方程式（Webster's hone equation）であらわされ，これを適当な境界条件の下でとくことができれば，音声波形を求めることができる．

$$A(x)\frac{\partial}{\partial x}\left(\frac{1}{A(x)}\cdot\frac{\partial U}{\partial x}\right)-\frac{1}{c^2}\left(\frac{\partial^2 U}{\partial t^2}\right)=0 \quad (1)$$

ここで $U=U(x,t)$ は声道内音声気流の体積流〔cm^3/sec〕，c は音速〔cm/sec〕，$A(x)$ は場所的に変化する声道の等価横断面積〔cm^2〕である．

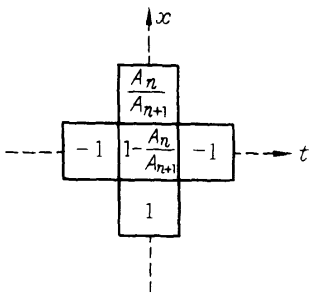
（1）式を，たとえば電子計算機によってとく最も直接的な方法は，これを差分の形に変換することであり，その結果は（2）式のようになる．

$$c^2(h'/h)^2\frac{1}{A_{n+1}}\{A_n(U_{n+1}-U_n)-A_{n+1}(U_n-U_{n-1})\}+2\phi_m-\phi_{m-1}-\phi_{m+1}=0 \quad (2)$$

ただし h: x の軸方向きざみ幅．n: その位置

h': t 軸方向のきざみ幅，m: その時点

ここで $h=0.8\times10^{-2}$〔m〕，$h'=1/4\times10^{-3}$〔sec〕，$c=320$〔m/sec〕とすれば $c^2(h'/h)^2=1$ となるから，（2）式は**第4図**に示すような差分演算子によって解かれる．しかし，この解法は $A_n/A_{n+1}>1$ のときは不安定となり発散してしまうため，一般的な解法として常に用いることはできない．



第4図 直接差分計算の演算子

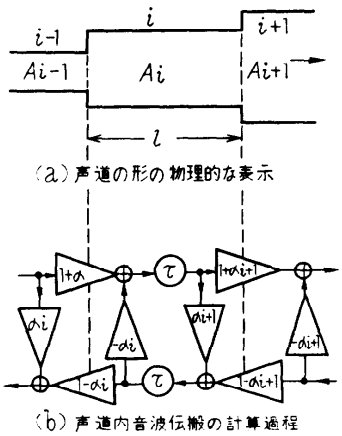
（2） 線路方程式による近似解

（1）式の波動方程式で $U(x,t)$ が $U(x)e^{st}$ のように変数分離できるとすれば，（3）式のようになる．

* 小さな単位から編集合成された音声の明瞭度と自然性は，当然他にくらべて劣っている．

** この近似は十分実用的な近似である．

$$\frac{d^2U}{dx^2}+\frac{1}{A(x)}\cdot\frac{dA(x)}{dx}\cdot\frac{dU}{dx}-\frac{s^2}{c^2}U=0 \quad (3)$$



第5図　不均一線路の微小均一区間接続による近似とその計算のフロー・チャート

これは不均一線線路の線路方程式にほかならない．そこで，この不均一線路とみなされた声道を，横断面積 $A(x)$ 一定の微小均一区間の接続として近似すれば，各微小区間はその長さによってきまる時間おくれと隣接区間との接続部における反射係数とでおきかえられ，結局**第5図**に示すような各区間ごとの差分計算で解を求めることができる．

この原理によって行なわれた BTL* での実例によれば[8]，

（a） $\Delta l=0.8$ cm の21区間（全長 16.8 cm）で近似する．

（b） 各区間の遅延時間 τ は 0.8 cm/320 m/sec＝1/40〔msec〕で，これを波形計算のサンプリング周期とする．

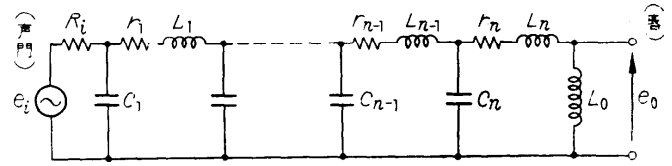
（c） 各区間の特性インピダンス Z_0 は $\rho c/A$ となり，断面積 A に反比例する（ρ は空気の密度）．

（d） 各区のつなぎ目を次のような反射係数で結ぶ．

$$\alpha_i=\frac{Z_0^{i+1}-Z_0^{i}}{Z_0^{i+1}+Z_0^{i}}=\frac{A_i-A_{i+1}}{A_i+A_{i+1}} \quad (4)$$

（e） 境界条件として，声間の等価面積を 0.2 cm²，唇からの放射空間の大きさを開口端面積の100倍とする．

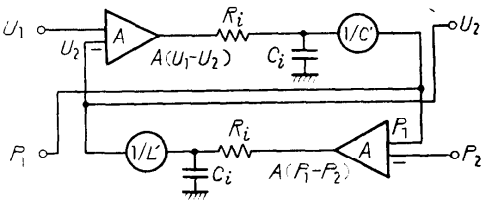
* ベル電話研究所



$$L_n=\frac{\rho}{A_n},\quad C_n=\frac{A_n}{\rho C^2}$$

（ρ: 空気の密度，C: 音速）

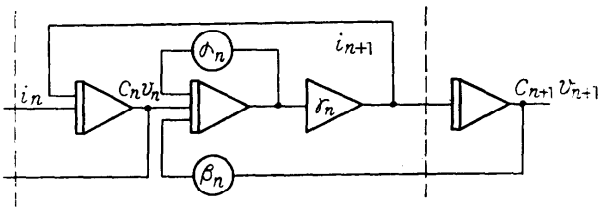
第6図　声道のL型集中定数表示



(a) MIT 方式

$$P_1=\frac{1}{C}\int(U_2-U_1)dt,\quad C=C'\frac{C_iR_i}{A}$$

$$U_2=\frac{1}{L}\int(P_1-P_2)dt,\quad L=L'\frac{C_iA_i}{A}$$



(b) 明大—松下通信機方式

$$a_n=\frac{\gamma_n}{L_n},\quad \beta=\frac{C_n}{C_{n+1}},\quad \gamma_n=\frac{L}{L_nC_n}$$

$$\begin{cases}\frac{1}{p}(i_n-i_{n+1})=C_nV_n\\ \frac{1}{p}(V_n-V_{n+1}-\gamma_ni_{n+1})\\ \quad=L_ni_{n+1}\end{cases}$$

第7図　L型集中定数表示のアナコン回路構成

この計算法は声道の模擬合成の計算機シミュレーション用としては非常にすぐれており，BTL では IBM 7090 をつかって合成実験を試みている．

このようなプログラムで連続音声を合成するためには次の 27 個のパラメターの値を時間の関数として与えなければならない．

声道各区間の等価横断面積	21 個
鼻腔と口腔の結合度	1
有声音源の強度とピッチ周波数	2
気音源の強度	1
無声音源の位置と強度	2

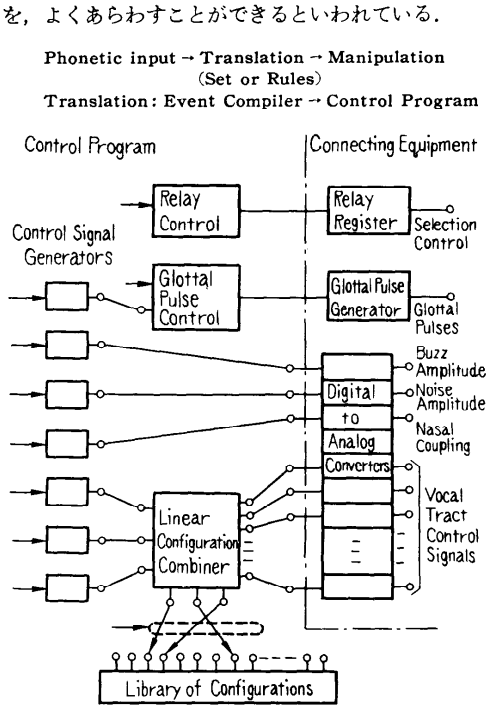
BTL の研究では，制御入力のカードに次の 6 個のパラメターを指定するようになっている．音韻の名称，母音のストレス，音源強度，ピッチ周波数，音韻間の過渡時間，音韻の継続時間．

制御の第 1 原則は，すべてのパラメターはその旧い値から新しい値にむかって時間的に linear に変化するように制御される．また声道パラメターの新しい値は，入力の音韻に対応して stored table の中から選び出される．

（3） アナログ計算機によるシミュレーション

第 5 図 (a) に示したような声道の物理的な近似は，音電変換の基本的な関係によって，**第 6 図**に示すような L 型集中常数回路の縦続接続によって，電気的にあらわされる[7]．このような等価回路はアナログ計算機によっていろいろに構成される．その実例をあげると**第 7 図**のようである．ここで (a) 図は米国 MIT での回路構成を示し[9]，(b) 図は明治大学と松下通信工業の共同研究によるものを示す[10]．

このようなアナログ計算回路の構成で連続音声を合成する場合，問題になるのはやはりその連続的な制御の方法である[7]．MIT の計画では**第 8 図**に示すような原理による計算機制御が考えられている[11]．ここで event compiler というのは，入力の音韻記号からそれに対応した音声を合成するために必要な events のリストを，合成法則にしたがって準備するものである．実験者はここでそのリストを変更したり，パラメターの値を変更したりすることができる．control program は event compiler の出力である events のリストから，実際の制御に必要な出力の形に時間的にその情報を組み立てる．声道の面積の変化の計算は，library からえらばれたいくつかの基本形の線型結合として行なわれる．しかも，この結合の係数を時間的に 2 次曲線的に制御する．このようにすれば library の内容を少なくして，しかも声道の形の時間的な変化を，よくあらわすことができるといわれている．



第 8 図 MIT における VT 型アナログ合成装置の制御方式

3.2. Terminal Analog 型音声合成

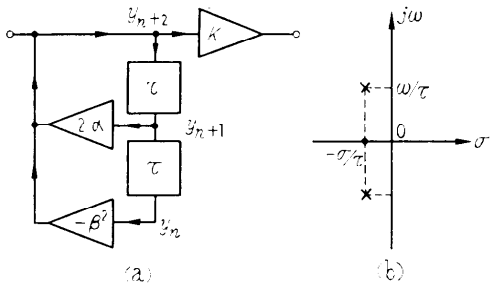
音声の調音器官（共鳴系）としての声道の特性は，その共振/反共振の周波数（この系の伝達関数の極と零点に対応する）で主として記述され，従来の音声研究の結果によれば，低次の 3～4 個の共振周波数と，1～2 個の反共振周波数が最も重要であり，音声の音韻情報の近似としてはそれで十分であることが知られている[7]．そこで，この共振/反共振特性によって声道の調音機能を現象的に代用させることができる．このような原理にもとづく音声合成の方法を terminal analog 型という．

（1） 計算機シミュレーション

声道の共振特性は複素共役極であらわされ[7]，一対の複素共役極は一般に 2 階の差分の形で表示することができる．その一番簡単な流れ図を**第 9 図**に示す[12]．ここで τ は時間おくれを示す．ここで

$$y_{n+2}-2\alpha y_{n+1}+\beta^2 y_n=0 \qquad (5)$$

という差分方程式が成り立つ．したがって（5）式の特性方程式 $\lambda^2-2\alpha\lambda+\beta^2=0$ の根は $\beta^2>\alpha^2$ のとき複



$y_{n+2}-2\alpha y_{n+1}+\beta^2 y_n=0$
$\beta=e^{-\sigma}$, $\alpha=e^{-\sigma}\cos\omega$, $K=1-2\alpha+\beta^2$
τ は sampling の unit delay

第9図　一組の複素共役極をあらわす Flow chart

素根となり，解は振動解

$$y(t)=\beta^t(A\sin\varphi t+B\cos\varphi t)$$

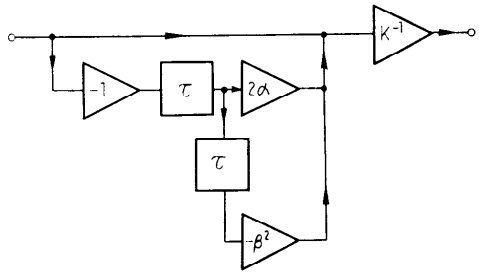
となる．ここでさらに $\beta=e^{-\sigma}$，$\alpha=e^{-\sigma}\cos\omega$ とすれば

$$y=e^{-\sigma t}(A\sin\omega t+B\cos\omega t)$$

となり，$t=0$ で $y=0$ という初期条件から $B=0$，$A=K$ となり，さらに定常状態における利得が1となるようにすれば

$$K=1-2\alpha+\beta^2=1-2e^{-\sigma}\cos\omega+e^{-2\sigma}$$

となり，すべての係数が求められる．結局第9図 (a) の流れ図で $\alpha=e^{-\sigma}\cos\omega$，$\beta=e^{-\sigma}$，$K=1-2e^{-\sigma}\cos\omega+e^{-2\sigma}$ とおけば，これは同図 (b) に示すような一対の共役極すなわち声道の共振特性(一つのホルマント)をあらわすことになる．反共振特性（零）は簡単な逆変換の関係を用いることによって，容易に極からみちびかれ**第 10 図**のようになる．



第10図　一組の共役零をあらわす Flow chart

この方法は，terminal analog 型の他の形での計算機シミュレーション[13,14]にくらべて，計算時間の上で非常にすぐれている．

（2） アナログ計算機シミュレーション

声道の特性の基本をなす単一の共振特性は，次のようにあらわせる．

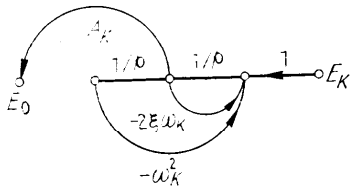
$$H_K(p)=\frac{A_K p}{p^2+2\xi\omega_K p+\omega_K{}^2(1+p^2)} \qquad (6)$$

ここで $\xi\omega_K=-\sigma_K$ がこの共振の damping を与える．

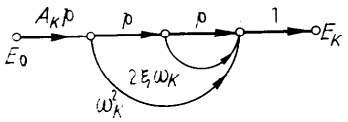
（6）式で関係ずけられる入力 E_0 と出力 E_K の関係は

$$E_K\approx\frac{(p^2+2\xi\omega_K p+\omega_K{}^2)}{A_K p}E_0 \quad ; p^2\ll 1 \qquad (7)$$

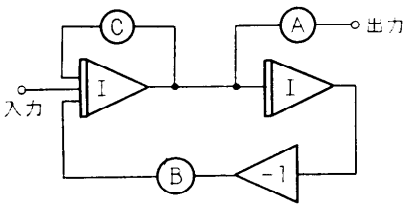
とかきかえられ，**第 11 図** (a) のような signal flaw graph であらわされる．したがって，その逆関係をあらわす（6）式は，第11図 (b) のような流れ図であらわされることになり，第11図 (c) のようなアナログ計算機の構成で計算されることになる[15]．ここで $A=A_K$ は K 番目のホルマントの振幅を，$B=\omega_K{}^2$ はその周波数の自乗を，$C=2\xi\omega_K=-2\sigma_K$ はその帯域幅をあらわす．



（a）（7）式の関係をあらわす Flow graph



（b）（6）式の関係をあらわす Flow graph



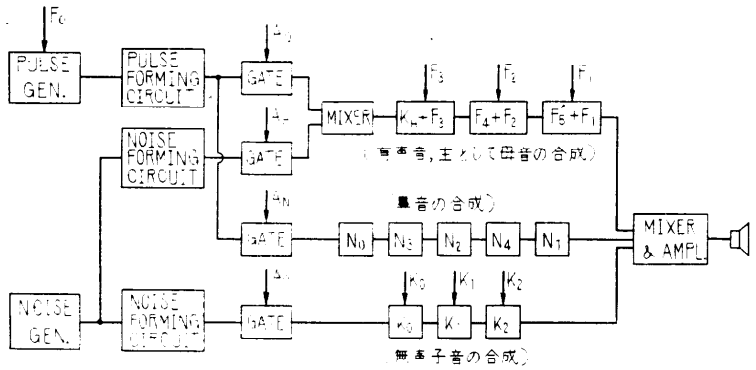
（c）（b）の flow graph をあらわすアナログ計算機の結線
$A=A_K$, $B=\omega_K{}^2$, $C=2\xi\omega_K$, I: inlegrator

第11図　アナログ計算機による共役極（ホルマント）の表現

東大の RAAG の音声研究グループではこのような原理によってアナログ計算機による音声合成（terminal analog の並列型）を実験している．

（3） 電気的アナログ回路による方法

この型のもっとも代表的なものがスエーデンの



第12図 OVEII のブロック図

RIT* の G. Fant 等の OVEII である．その構成のブロック図を**第12図**に示す．Fant 等はこの装置を用いて限られた例ではあるが，1,000 bits/sec の制御情報で人間の音声とほとんど区別のつかない合成音声を作った．また，このような方法では real time で音声合成ができる点はなはだ便利であり，また実用的である．なお，この電気的なアナログ装置についてはくわしい論文[16]があるので，これ以上のことは省略する．

4. 連続音声合成の問題点

2, 3 節に実際に音声波形を合成する各種の方法について，できるだけ系統的に紹介してきた．しかし最初にのべたような目的，すなわち文字や音韻記号のような不連続な入力から連続音声を合成するという問題で一番の難問は，このような合成の手段についてではなくて，やはり音声そのものの性質に関した問題，いわゆる“langnage problem”である[17,18]．

Langnage problem というのは，文字や音韻記号のような音韻的（phonemic）な情報のみから，まず発音単位としての word を構成し，その word の言語的な働きから，それが連続音声としてどのように発音されるかをきめ，その発音に適した合成装置の制御を行なうという問題である．この問題は音声そのものの特性，本質に関する問題であるので，ここではこれ以上立ち入らないことにするが，現段階における音声合成の主要問題として，内外で盛んに研究，実験が行なわれている．

5. 計算機による合成実験の例

我々のところでも，計算機による computer simulation の形で音声合成の実験的研究を行なっているので，その内容を簡単に示そう．

5.1. Terminal Analog 型

3.2. 節の（1）でのべた2階の差分計算による方法で，三つのホルマントを持つ音声の合成実験（主として母音）を，計算機 NEAC-2206 を用いて行なった．実験の結果は（当然のことながら）ホルマント周波数の制御が適切であれば，十分明瞭な良質の合成音声がえられる[19]．制御入力としては次のものを与える．

（1）ピッチ周波数 f_0,（2）音源強度 I,（3）音源波形定数（τ_1 と K),（4）第1から第3までのホルマント周波数 f_i とその帯域幅 B_i ($i=1, 2, 3$).

NEAC-2206 による計算で time scale は 2×10^3 程度（1秒間の音声の合成が約30分）であった．

5.2. Vocal Tract Analog 型

3.1. 節の（2）でのべた線路方程式による近似解（BTL 方式）による方法で実験を行なっている．詳細については別に論文として発表の機会をえたいと思っているので，ここにはのべないが，計算方法として問題は time scale が NEAC-2206 で 1.3×10^4 程度かかり，約1秒間の音声を合成するのに3時間以上かかることである．したがって，合成のための制御法則に対して，我々の設定した諸仮定が適切であるかどうかを．まだ十分検討し尽すには至っていないが，現在のところ音韻記号で書かれた入力を主とし，その各音韻の継続時間と主として有声音源の特性を時間的に規定する情報（たとえばピッチ周波数とか音源強度など）を副次的に与えることによって，一応それらしく聞える連続音声を合成することができるといった段階である[20]．

6. むすび

最後に音声の機械による自動識別の問題と音声合成の研究の関係について簡単にふれておく．

音声が人間によってどのような仕方で認識されるか，という perception model の問題については，今なお不明のことが多く憶測の域を出ない．しかし，それらの中で現在のところ比較的合理的と考えられるものに，Articulatory Reference Theory とよばれ

* Royal Institute of Technology

るものがある[21]．簡単にいって，音声を聞いた人が同じ音声を自分がどのようにして発声するか，という調音運動（狭義の articulation）の感覚を仲介として音声認識が行なわれると考えるもので，その実現の工学的な手段として，Analysis-by-Synthesis (or Active Analysis) という方法が提案されている[21]．これは内部に合成モデルを内蔵した比較系を考え，入力と合成出力の error が最小となるように内部合成の情報パラメターを制御し，その最適な制御パラメターの値として入力の情報を抽出する方法である．

このような原理と方法によって音声の識別を行なおうとすれば，直接音声波形を合成する必要はなくとも，そのために必要なスペクトル特性とか，声道の形の情報とかいうものを最終的な識別単位としての音韻とか，音節あるいは単語といった不連続な単位から合成する過程が明らかにされなければならず，そのために単に定性的な説明的なもののみでなくて，数値的に operational な合成の法則を明らかにしなければならない．このような意味から，現在では音声の coding mechanism としての合成の研究と decoding process としての識別の研究は表裏一体をなすものと考えられている．

はしがきにおいてのべたような実用的な音声合成の目的には，どのような原理，方式が最適であろうか．

多くの合成法が考えられてきたが，その一つの typical な例が「法則による合成」であり，記憶容量は少なくてよいが，合成の法則を簡単に例外的な修飾を少なくすればするほど，人間の発声機構とのアナロジーの程度の高い高級な音声合成装置と複雑なロジックとを必要とする．しかし spelling の問題は起らない．これと対照的なのが編集による合成であり，アナログ的な音声合成装置は必要としないが，大容量の random access の記憶装置を必要とし，spelling の問題をさけることはできない．結局記憶容量の大きさと合成，制御の高級，複雑さとがお互に取り引きされているようなものである．

この二つの typical な方法の中間にいくつかの中間的な hybrid system が考えられる．たとえば，入力の文字を音韻的な記述に変換するのに，発音に必要な情報まで書きこんだ dictionary をさがし，その結果によって合成装置をある法則にしたがって制御する．または入力情報から合成装置を制御する信号を直接制御電圧の形で stored table からよみだす．

このような多くの可能性の中で，装置（記憶，制御，合成）が最も経済的で，しかも高品質の連続音声を合成しうる可能性が多いものとして phonemic dictionary look-up と synthesis by rules を組み合わせたものが最も有利なものであろうといわれている[6]．しかし実用的な意味で合成する vocabulary に適当な制限を加えられるような場合には，編集による合成も十分有用なものといえよう．

電話による音声の伝送ということからおこってきた音声の工学的な研究も，音声波形の伝送という問題から音声の情報の伝送という問題にすすむにつれて，音声の聞き取り（perception）という問題に立ち入らざるをえなくなり，今日では分析，合成，伝送，認識のあらゆる面で情報処理的な色彩の濃い研究が必要となり，また事実多くなったと思われる．また音声の認識や合成ということはある意味で，将来の計算機や他の情報処理装置の姿を考えるとき，直接に，間接に何らかの考慮を払わなければならない問題となりつつあるように思われる．この機会に情報処理研究の専門の方々に音声研究への今後の御理解と御協力をお願してこの解説を終らしていただく．

参考文献

1) Hariss C.M.: A Study of the Building Blocks of Speech. JASA, **25**, 962～969 (1953).
2) Inomata, S.: A New Scheme for Speech Generation, S.S.S. 電気試衆, **24**, 47～57 (1960).
3) Lee L.H. and Mulvany R.B.: Now a Talking Computer Answers Inventory Inquiries. Electronics, Aug. **16**, 30～32 (1963).
4) 橋本　清，関口　茂：電話番号案内用音声の編集試聴実験，日本音響学会誌，**20**, 241～248 (39 年 7 月).
5) Sivertsen E. and Peterson G.E.: Studies on Speech Synthesis. Rep. No. 5, Speech Res. Lab., Univ. Michigan (1960).
6) Cooper F.S.; Speech from Stored Data. IEEE. Conv. Paper 53.2 (1963).
7) Fant G.: Acoustic Theory of Speech Production- Mouton & Co. s'-Granenhage (1960).
8) Kelly J.L. Jr. and Lockbaum C.: Speech Synthesis. Speech Comm. Seminar., Stockholm (1962).
9) Whitman E.C.: Transistorized Articulatory Speech Synthesizer. QPR, MIT-RLE., No. 68, 164～167 (1963).
10) 吉田登美男，小川康男他：ボーカルトラクトのアナログコンピュータによるシミュレーション，

音学会音声研究会資料（39 年 4 月）.
11) Dennis J.B.: Computer Control of an Analog Vocal Tract. Speech Comm. Seminar, Stockholm (1962).
12) Flanagan J.L. et al.: Digital Computer Simulation of a Formant-Vocoder Speech Synthesizer. 15 th Anual Meeting of Audio Eng. Soci. No. 307 (1963).
13) 猪股修二：電子計算機による音声の発生について，日本音響学会誌，**17**, 93～102 (36 年 6 月).
14) 橋本新一郎，松本光晴：計算機による母音の合成，日本音響学会誌，**20**, 385～395 (39 年 11 月).
15) The RAAG Phonetical Group: Some Preliminary Experiments on the Varification of Principle of the Composition and Decomposition of Phonemes. RAAG Memoirs, 111, H, 693～713 (1962).
16) Fant G. and Mártony J.: Instrumentation for parametric synthesis (OVEII). STL-QPSR-2/1962, 18～24 (April-June. 1962).
17) Cooper F.S. et al.: Speech Synthesis by Rules. Speech Comm. Seminar, Stockholm (1962).
18) Peterson G.E. and Fillmore C.J: The Theory of Phonemic Analysis. Internalt. Conf. Phonetics., Helsink (1962).
19) 光岡輝義，中田和男，平松啓二：音声の計算機による合成（1），音学会研究発表論文 1-3-6 (39 年 5 月).
20) 光岡輝義，中田和男，平松啓二：声道の形の情報による音声波形の合成，音学会音声研究会資料（40 年 3 月）.
21) Stevens K.N.; Toward a Model for Speech Recognition. JASA, **32**, 47～51 (Jan. 1960).

（昭和 40 年 3 月 29 日受付）