

音声応答と音声合成

中 田 和 男

(日立製作所 中央研究所)

1. 音声応答

計算機の利用形態が発展するにしたがって、とくにオンラインでの利用がすむにつれて、マン・マシン・インターフェースの1つとして音声による情報の入出力が注目されるようになってきた。

しかし音声による情報の入力、ユーザーの期待にかかわらずその技術的な困難さゆえに、ごく限られた例を除いては実験室レベルの研究にとどまっており、実用化までにはいたっていない。

これにくらべて音声による情報の出力は、電話による時報サービス、簡易計算サービス、座席予約サービスなど数多くの具体例がすでに実用に供されている。

入力情報によるリクエストに答えて情報処理の結果を音声の形でユーザーに出力するための装置を音声応答装置 (Audio Response Equipment or Voice Answer Back System) と定義すると、その音声出力方式はその要求される内容によって、各種存在する。その方式を決定する最大の要因は、出力音声に要求される異なる語彙 (以後これを語い (vocabulary) という) であり、同時応答可能な回線数 (多重度)、音声品質、コストなどが他の要因である。

いま音声応答をその語い数によって分類してみると、

- (1) 少数限定語い —— 時報サービス、計算サービスなど
- (2) 多数固定文章 —— 駅における列車案内、空港の航空線発着案内など
- (3) 多数可変語い —— 電話座席予約、株価案内など
- (4) 無限可変名詞 —— 電話番号案内、銀行通知サービス
- (5) 無限語い文章 —— 汎用音声応答

時報サービスや計算サービスでは、応答語い1桁もしくは2桁の数字音と「時」、「分」、「た」、「いまの時刻は」、「です」、「あなたのお答は」などの少数の句であり、その総数は50~100語程度である。

駅や空港における列車や航空線の発着案内では、時刻、行先、始発点、ホーム (ゲート) 番号、(航空会社名) などかなり複雑な内容をもつが、その数は500語程度である。列車の場合、普通はその情報を必要とする順序があらかじめ定まっており、変更を必要とする場合は人間がアナウンスするとすれば、シーケンシャルなアクセスメモリー (たとえば磁気テープ) でよいから比較的簡単になる。

電話座席予約では、日時、時刻、駅名、列車番号、列車愛称名、座席の種類、人数、金額など語い数2000語以上に達する。

各種の通知・案内サービスでは、応答文のしくみはさまっているが、その内の番号、人名、地名、会社名などのキー・ワードが可変であり、その語いは無制限と考へなければならぬ。

汎用の音声応答は、原理的には文字で表記されるすべての言語情報を音声という形にかえようとするもっとも野心的な試みであるが、現在なおまだ技術的に解決しなければならぬ多くの問題をかかっている。

もう一つ、日本における一般ユーザー (末端ユーザー) サービス用の音声応

答に、ユーザー（中央ユーザー）が要求する1つの特色は「女性の声で」ということである。経験者にきくと、アメリカなどでは、専用ユーザーを主対象としているせいかすべて「男性の声」で実用化されているそうであり、面白い対比を示している。しかしあとで説明するよう「女性の声」を作ることは、単純な録音・再生方式をのぞけば、「男性の声」を作ることより技術的にむつかしく、メーカー泣かせと現状でなっている。

実用化の観点からみれば、この10年間のうちに、(1)、(2)、(3)のレベルまですでに実用化されており、現在(4)のレベルが実用化に近づいているといえる。

(5)のレベルについてはまだ実験室レベルに止まっているといっている。

2. 音声合成技術

音声応答（装置）を実現するには、録音・編集をも含めた広い意味での音声合成技術を利用している。その内容をごく簡単に説明する。

音声合成技術は次のように分類することができる。

(a) 録音・編集方式

(b) 録音（音声）素片連結・編集方式

(c) 合成方式

(1) 分析・合成方式

(イ) パラメータ編集方式

(ロ) パラメータ連結・編集方式

(ハ) 規則合成方式

録音・編集方式というのは、よく知られているように、ある単位たとえば単語、句などで必要とする語いをあらかじめ録音しておき、応答に際して必要とする語を選択し編集して出力するものである。録音の単位としては言語として意味をもつ単語もしくは句のレベルが使われるのが普通である。

技術的な問題点としては、記憶媒体の容量の有効な利用のため単語の時間長をある単位、整数倍の時間の中に過不足なく記録すること、出力音声の品質を高めるために、たとえば数字音のようなものは何種類かの變形（主として声の調子の変化によるもの）を記録しておくことが必要なことである。

録音・編集方式で応答しうる語いをふやさうとすると比例的に必要記憶容量が増大し装置として高価なものになる。それを解決するには2つの方法がある。

1つは音声合成技術（狭義）を利用し、情報をパラメータ（音声合成装置を動かすための制御情報）という形で記録することによって情報圧縮を行う方法であり、これが(c)の合成方式である。

いま1つは録音編集における記憶単位を言語の表現として意味をもたない程小さいが音声として1つの単位と考えられる大きさにするのである。

たとえば日本語の場合、音節（正しくはモーラ：拍）という単位で考えてみると、言語音声的にはすべての日本語音声は1/3種の音節の組み合わせ（連結）でありぬさぬ。粗くいえば、カナ文字1字に相当する音を単位として必要な種類数だけ（たかだか1/2の程度）記憶しておき、そのつなぎ合せで単語や句を作り出すと考える。具体的には「音声」という単語を、単独に1つづつ発音された「オ」、「ン」、「セ」、「イ」という4つの音節のつなぎ合せで作るわけである。

もしこのようにして作り出された音声がい分明瞭で自然であるならば話は簡単である。しかし実際にけ意味を理解することはできてはけなけだ不自然な音声しか作り出せない。この原理に自然にきこえるような改良を加えたりが (b) の録音(音声)素片の連結・編集方式である。

ちなみに、音声波形をデジタル記憶するとして必要情報量を概算してみると次のようになる。

サンプリング周波数: 8 kHz 量子化ビット数: 8 ビット
 1 単語の平均時間長: 1 秒 必要語の数: 2000 語
 情報量 = $8 \times 8 \times 10^3 \times 1 \times 2 \times 10^3 = 128 \times 10^6$ ビット = 128 M ビット = 16 M バイト
 一方 120 コの単音節の記憶に必要な情報量は、平均音節長を 0.2 秒として、
 $8 \times 8 \times 10^3 \times 0.2 \times 120 = 15.36 \times 10^5 \approx 1.6$ M ビット = 200 名バイト となる。

技術的な問題点はいかにして出力音声の品質とくに自然さを録音・編集方式に近づけるかである。

1 コのまとまった単語として発声された「音声」と、単独に発声された単音節の単純な連結による「オ・ン・セ・イ」とをその物理的な特徴で比較してみると、

物理特性	単語	単音節の単純連結
1) ピッチ	固有のまとまりの形	単音節内で固有のまとまり、境界で不連続(音節間で独立)
2) 音節時間長	単語内での音節として固有の長さ	単音節として固有の長さ、すべての単語において同一音節の時間長は一定
3) ホルムント周波数	単語内では連続で固有のパタン	単音節内で連続で固有、境界で不連続(音節間で独立)
4) 強度(振幅)	単語として固有の形(連続的な包絡)	単音節内で固有の形、音節間で独立(音節の境界で不連続)

この連続と不連続(音節間で独立)の対立のなかで、どの要因が自然さに対してもっとも重要な聴覚的にしよべ、その重要性に応じて近似的にでも連続化する処理が必要である。

我々の行った実験の結果でけ、上表の 1), 2), 3), 4) の順に単語としての自然さへの貢献が大きいことがわかった。したがって問題け、あらかじめ録音されている単音節単位の音声波形を必要な順にえし出して連結するにあたって、そのピッチと時間長をいかにして制御するかということになる。

音声波形の我々に与えるピッチ感覚(それによつて単語にアクセントが付き、句や文にイントネーションが付与される)け、音声波形のこまかくりかえし性に起因するもりであり、このくりかえし性け、(有声音の)音源としての声帯振動(波形)のくりかえし性によるもりである。

たとえげ女性声「ア」の波形けその例を国 1 に示すようなもりであり、明らかに波形のくりかえしを観測することができける。このくりかえしの 1 周期分の波形は、1 回の声帯振動による共鳴系としての声道(声内から唇までの調音空間)の応答出力、単純化してえげ 1 コの音源インパルスによる声道のインパルス応答波形、と考えることができける。

そこで 1 音節の音声波形をあらかじめこのピッチ周期ごとに区分して、国 2 に示すように記録しておき、読出しにあつてこの波形を制御情報によつて指定された時間長だけ読出すということにすれば、原理的に再生音声のピッチを制御することができける。

またその総時間長をも可変にすることができ、2) の音節時間長の制御も可能になる。このような方式がいわゆる「ピッチ単位音声素片による音声合成」(または「音声素片のピッチ同期編集合成方式」)とよばれる方法である。⁴⁾

以上 (a), (b) は音声波形を作り出すにあたってなんらかの形で人間の音声波形を直接的に音声波形素材として利用している。これに対して (c) の音声合成方式は、人間の音声波形を直接的に合成の素材として利用するということとせず、一たん制御情報(パラメータ)に変換し、音声合成装置によって音声波形に再生する方式といえることができる。

パラメータのえらび方は音声合成装置の方式に対応しており、現在までのところ主要なものとして次のようなものが提案され実験されている。

- i) 周波数スペクトルエンベロープ — Vocoder 型
- ii) ホルメント周波数 — Terminal Analog 型
- iii) 声道面積関数 — Vocal Tract Analog 型
- iv) 線形予測係数(LPC) — 線形予測フィルター
- v) 偏相関係数(PARCOR) — PARCOR フィルター
- vi) 波形情報圧縮方式 — DPCM, ADPCM など

ア) の分析・合成方式というのはいわゆる情報圧縮伝送方式と同じであり、たゞその分析結果を一たん記憶し、いつでも、何回でも再生できるようにしたものである。

イ) のパラメータ編集方式というのはいわゆる録音・編集と原理的には同じで、たゞ制御情報という形で記憶し、必要記憶容量の低減をはかっている。

ウ) のパラメータ連結・編集方式というのはいわゆる (b) の方式をパラメータの次元で実行しようというもので、波形次元での (b) 方式にくらべて、ホルメント周波数や振幅をもスムーズに連続しようという利点がある。

エ) の規則合成というのはいわゆる音声の出力に必要な制御情報そのものをも、できるだけあらかじめ入力情報と合成規則(制御情報生成規則)とから作り出し、それとをさしに音声波形に変換しようというもっとも野心的な試みである。⁵⁾

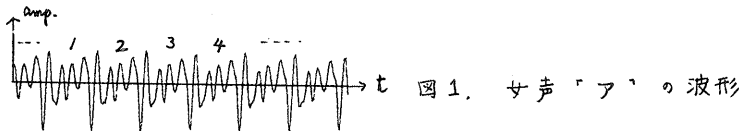


図1. 女声「ア」の波形

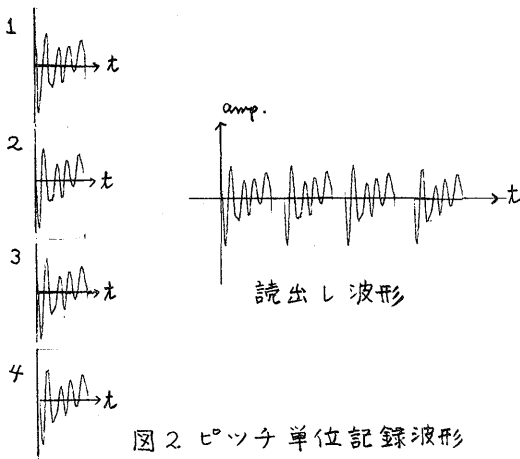


図2. ピッチ単位記録波形

3. 音声応答と音声合成技術

従来実用化されてきた音声応答装置はすべて分類の(1)、(2)の段階のものであり、合成技術としてけずべて録音編集方式でまかなわれてきた。

録音編集方式には、出力音声の品質が高く、技術が簡単で多重化も容易である、といった利点があり、他の方式は出力音声の品質の点でなかなかおよばない。

ことに最近では計算機およびその周辺技術の急速な進歩で、大容量ファイルとして大型の磁気ドラムやディスクが発売され、ついに最近の国鉄の電話座席予約システムでは(3)のレベルまで録音編集方式で実現されてしまった。

現在技術的にけ(4)のレベルが(b)の方式で実用化される寸前まできているが、なお出力音声の品質の点で必ずしも十分実用(ことに不特定多数の一般ユーザを相手としたとき)にたえうると確言できるにはいたっていない。

(b)方式をこのまで実用化に近づけた最大の原因は、ピッチ単位音声素片波形を自然音声波形の区別から作るという方法と、PARCOR(またはこれと等価なLPC)分析の結果によるPARCOR合成フィルタのインパルス応答でおきかえるという方法とをつたことであるが、現在合成音声の品質が最後のところで十分に確保されないという原因もまたPARCORインパルス応答を使っているというところにあると思われ(この点については4.で詳しく説明する)。

米国においても音節もしくはそれに類似のCVC(日本ではVCV)または単語(数字音)の連結による句、文の合成を実用を目的に試みているが、英語においてはその音韻的な構造やアクセント要因のちがいが波形次元における(b)方式ではなくてパラメータ次元における(c)の(ウ)方式を主力に研究している。

規則合成方式はまだその内容について多くのレベルで研究の段階にあり実用にはいたっていない。

我々もこの数年方式(b)を主に研究をつづけてきた³⁾。その背景には、

- ・ 音声出力は結局はメモリー(波形または制御情報)と処理(制御情報の作成および音声波形合成)のtrade offであり、将来はメモリーを多く使い処理を少くする方式の方が技術的にもコスト的にも有利となるのではないが、
- ・ 自然音声の内に自然に内ぞうされてあり、合成的にはまだ我々が作り出すことのできないなにかunknownな自然さ(人回りの声、ことに女性の声らしさ)を利用することができないのではないが、
- ・ 単音節単位で制御できるとなると音韻的な内容の制御は非常に簡単になりうるのでないか
- ・ PARCORフィルタのインパルス応答という形でピッチ単位音声素片を作ることをできるとすれば、あらかじめ記憶しておくべき音声素片を、対応する肉声の単音節波形からほとんど自動的に作り出していくことができる、といった期待があったからである。

現状の合成音声のサンプルを録音で示す。

内容は女性の声による下記の文章である。

「こちらは、日立銀行コンピュータセンターで ございます」

「まいど ありがとうございます」

その他

4. 音声合成実用上の向題点

録音編集以外の音声合成技術が音声応答装置として実用化されない最大の原因は、出力音声の品質が十分でないことである。したがって音声合成実用化のための向題点けつぎり2つに集約される。

- 1) 合成音声品質とくに自然さ、人間の声(女性の声)らしさの改善
- 2) 合成音声の品質評価法の確立

音声合成技術の現状では合成音声の明瞭さ(意味内容の理解)はまず確保されているといつてよい。向題は自然さとくに女性の声らしさの改善である。

一方ある音声合成技術がある発展段階に達したとき、それが実用化にたえうる品質の音声と出力できるか否か判定が向題になる。しかし現在合成音声の品質と明瞭さと自然さの両面にあつて客観的に評価しうる手段がないので、その判定は個々の個人の主観的な評価にたよらざるをえない。そこで混乱と不信がおこる結果となる。

4.1 合成音声の品質の改善とくにPARCOR方式を利用する場合

1) PARCOR分析法の本質

PARCOR分析・合成法は、板倉・斎藤⁷⁾によつて発表されて以来数多くの実験によつて音声の分析・合成アルゴリズムとしての有効さが立証されてきた。

一方理論的にその本質についての研究もすゝめられ、いまや次のような理解が共通の理解として確立されてきた。

a) 音声の発生機構を次の偏微分方程式にしたがう声道中の音波の伝播と考へる限り、等価的な線形システムと考へてよく、その伝達関数は極のみでなく零をも有する

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

$U(x, t)$ 音波の体積速度
 $A(x)$ 声道の横断面積
 c 音波の伝播速度

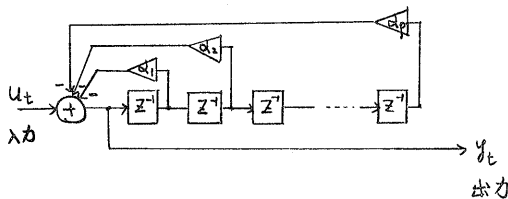
$$H(z) = \frac{F(z)}{G(z)} = \frac{1 + \beta_1 z^{-1} + \dots + \beta_p z^{-p}}{1 + \alpha_1 z^{-1} + \dots + \alpha_p z^{-p}}$$

伝達関数のz表示

b) PARCOR分析の基礎になつてゐる線形予測分析(LPC分析)は、上記のシステムをall-pole(自己回帰)モデルと仮定したときの、観測出力による線形システム同定手法そのものである。

$$H(z) \approx \frac{1}{G(z)} = \frac{1}{1 + \alpha_1 z^{-1} + \dots + \alpha_p z^{-p}} \iff y_t + \sum_{i=1}^p \alpha_i y_{t-i} = u_t$$

$\{\alpha_i\}$ 線形予測係数



システムの同定は次のYule Walkerの方程式を解くことによつて行なわれる。

$$\begin{bmatrix} f_{11} & f_{12} & \dots & f_{1p} \\ f_{21} & f_{22} & \dots & f_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ f_{p1} & f_{p2} & \dots & f_{pp} \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} = - \begin{bmatrix} f_{10} \\ f_{20} \\ \vdots \\ f_{p0} \end{bmatrix}$$

$(x_0, x_1, \dots, x_{n-1}, x_n)$ 音声波形

$$f_{ik} = \sum_{n=0}^{N-1} x_{n-i} \cdot x_{n-k} = f_{ki} \quad (\text{共分散})$$

2) PARCOR分析法の限界と拡張

PARCOR分析は次の2つの仮定のうえに成り立っている。

a) システムは all-pole (自己回帰) モデルで近似できる

b) 入力は白色ガウス雑音である

これに対して現実の音声生成モデルは、

a) pole-zero (自己回帰・移動平均混合) モデルである

b) 音声波形の大部分をしめる有声音区間では、入力は(等価的に)インパルスの系列(series)である

このモデルと現実のくいちがいによって、分析抽出されるパラメータに次の原理による誤差が含まれる。

1) ピッチ周波数の効果 —— とくにピッチ周波数が高く線スペクトル密度の粗い女声においては、真の声道共振としてのホルマント周波数ではなくて、それにもっとも近いピッチ高調波周波数をホルマント周波数として与え、ホルマント帯域幅も極端に狭い値となる。

2) 零周波数の効果 —— 等価的な極フィルタの特性を与える。

$$H(z) = \frac{1 + \beta_1 z^{-1} + \dots + \beta_L z^{-L}}{1 + \alpha_1 z^{-1} + \dots + \alpha_p z^{-p}} \approx \frac{1}{1 + \alpha_1 z^{-1} + \dots + \alpha_p z^{-p}}$$

分析・合成系として、分析時の入力波形を再生する目的で PARCOR 方式を用いるときは、分析時のピッチ周波数とほとんど同じピッチ周波数で音声を合成するから上記のパラメータ抽出の不正確さは補正されて余り顕著にけありぬれない。

しかしピッチ単位の音声素片の作成に使うときは、その時々に応じて分析時とは異なるピッチ周波数で再生されるからその効果は無視できない。

線形システム固定理論と零特性を含む場合にも拡張する試みは提案されているが、鼻音のような共鳴系の形状による zero の抽出に注目しており、音源波形による zero には注目されていない。

線形システム固定理論と音源がインパルス系列であり、しかも出力が観測できないときに拡張する試みもなされているがまだ十分なものではない。

ことに zero を含む系で音源がインパルス系列であるという組み合わせに対して有効な分析法はまだ確立されていない。

3) 残差波形の利用

PARCOR方式では分析の結果、いわゆる残差波形とその周期性を目安としてインパルス系列か白色ランダム雑音ができかえて単純化し、大幅の情報圧縮(9,600ビット/秒程度)を行っている。しかし反面品質を劣化させていることはいふまでもない。そこでこの残差波形をなんらかの方法で有効に情報圧縮してそのまゝ記録し、再生時に復元して音源として利用する方法が考えられる。しかし残差波形は非常に複雑な情報をもっており、簡単な方法では余り効果はあがっていない。

ピッチ単位素片波形の作成にあつても、残差波形を少くともスペクトル上では保存するよゝな仕方で音源インパルスとして利用する方法が提案されている。¹¹⁾

4.2 合成音声の品質の改善(一般的な場合)

現在の音声合成技術で無視しているもの、それは1ピッチ周期内での特性の時間的な変化。ことに音源と共鳴系の結合による相互作用および声道壁面の放射効果などである。最近のベル研究所での石坂, Flanagan¹²⁾によるシミュレーション実験およびそれを聞いた人の報告によると、上記の効果も考慮した音源と声道モデルによる合成実験によつて大へん自然な人間らしい音声を合成できたということである。将来有望な一つの方法といふことができよう。

4.3 制御情報の作成

録音・編集方式以外の方法ではすべて音声的な制御情報が必要である。

分析・合成方式では分析法さえ自動的に確定したものであれば余り問題は無い。その他の方式では一旦肉声から抽出された情報を何らかのレベルで規則として抽象化し組織化することが必要である。たとえば単語の内容と、その音節数、アクセント位置(またはアクセント型)が与えられたとき、ピッチや音節時間長などのように制御するかを決定しなければならぬ。対象を人名とか会社名に限ってもなかなかむづかしい。単に多数のデータからの統計的な抽象だけでなく理論にもとづく解析とその結果の合成による検証が必要である。

4.4 男声と女声の変換

最近男声と女声のちがいを分析し、男声を女声に変換する研究が行なわれている。¹³⁾ それにはつぎの2つの理由がある。

1) 男声は女声にくらべてピッチ周波数が低く、ピッチ区間ごとの波形を互に独立と考へても近似がよいので分析が正確にでき、合成音声の品質も高い。そこで男声を変換して品質のよい女声を作ることはできないか。

2) 情報率内などの音声応答では女声が要求されることが多いが、場合によつては男声の方が望ましいこともある。そうしたとき、男声用、女声用と別々の制御情報をもつてではなくて、変換規則によつて一方から他方を作り出すことはできないか。

一般に女声は男声より合成しにくく、世界中のほとんどの音声合成は男声で試みられている。本当に女声らしい美しい合成音はまだ作られていないといつてよい。それはなぜか、それをつきつめていけば合成音声の自然さを改善する要因をみつけることができるかもしれない。

5. 音声合成の今後

音声合成研究の今後には次の3つの方向があるといえよう。

1) 音声応答装置への実用化をすすめる研究——合成音声の品質の改善とシステムとしての performance/cost との要求にはさまれながら地道に研究と実用化の gap をつめていく努力が必要であり、それを通じて technical know how が蓄積されていく。

2) ソフト的には規則合成を正統的におしすすめ、ハード的には(4.2のよう)の観点をとり入れながら音声合成研究の本質を解明していく。

3) 音声理解系 (speech understanding system) の研究と結合し、¹⁴⁾ Q/A過程を内包する「音声による質問応答システム」へと発展し、man-machine interface として新しい応用面を開拓していく。¹⁵⁾

さいわいに日本における音声合成研究のレベルは世界的で、ことに品質にうるさい日本人気質を反映して品質の高い研究が多く、1) はメーカーを主として、2) は大学を中心として、3) は官公社研究所を先導として研究がすすめられており、PARCOR方式の発明や電話座席予約のように研究面でも実用面でも世界をリードしているといえてよい。

今後とも我々はこの優位を保つべくより努力していきたい。

参考文献

- 1) 斎藤牧三；最近の音声研究について，電子通信学会研究会 EA75-26 (1975)
- 2) S. Sagawa, E. Shikiba, A. Suehiro et al ; Automatic Seat Reservation by Touch-Tone Telephone, 2nd USA-Japan Computer Conference, 14-2 (1975)
- 3) 市川, 中田；音声素片を用いた単音節編集型音声合成における音声素片の作成法, 電子通信学会論文誌 D, (印刷中)
- 4) a) 榎松, 井上；ピッチ単位音声素片の録音編集による音声合成シミュレーション, 日本音響学会研究発表, 2-1-2 (1970年10月)
b) 松井英一；音声素片のピッチ同期編集による多重化音声出力装置, 電気学会連合大会, 2457 (昭和43年)
- 5) a) J. L. Flanagan et al ; Synthetic Voice for Computers
IEEE Spectrum, (Oct, 1970)
b) E. J. Franzak ; The University of Michigan Audio Response System and Speech Synthesis Facility, 2nd USA-Japan Computer Conference, 18-3 (1975)
- 6) A. E. Rosenberg et al ; Speech Synthesis by Concatenation of Formant-Coded Words, JASA, 50-6, 1532, (1971)
- 7) 板倉, 斎藤；統計的手法による音声スペクトル密度とホルマント周波数の推定
電子通信学会論文誌 53-A, No.1, (昭和45年)
- 8) a) 松井英一；調音の最適推定の展望, 音響学会音声研究会, S75-08, (1975)
b) J. Makhoul ; Linear Prediction: A Tutorial Review, Proc. IEEE, 63-4 (1975)
- 9) a) 音声研究 '73 & '74, ETL-S9-1, 電総研パタン情報部音声認識研究室 (1974)
b) 深林, 鈴木；線形予測的音声分析の極・零型逆フィルタ法, 電子通信学会電気音響研究会, EA74-57, (1975)
c) 森川博由；最小二乗法を用いた極および零点を含む音声分析, 日本音響学会研究発表, 1-3-14, (昭和48年10月)
- 10) 河原英紀 他；線形予測を用いた声道伝達関数推定の一方法, 電子通信学会技術研究報告 Vol5 No.77, EA75-28 (1975)
- 11) 伏木田, 落合；素片編集型音声合成を目的とした自動分析・合成の一方式, 音響学会音声研究会 S74-23 (1974)
- 12) J. L. Flanagan, K. Iizuka ; Synthesis of Speech From a Dynamic Model of the Vocal Cords and Vocal Tract, BSTJ, 54-3, (1975)

- 13) a) 佐藤大和; 男女声の声質情報を決める要素, 電々公社通研研究実用化報告, 24-5, (1975)
- b) 安広, 尾関; 男声女声変換実験, 信学技報 Vol 75 No. 77, E75-29 (1975)
- 14) Speech Understanding System, Proc. IEEE Symp. April 1974
- 15) a) 箱田, 佐藤; 音声による質問回答システムの音声応答部の検討, 日本音響学会研究発表, 1.4.2. (1975年10月予定)
- b) 好田, 中津他; 音声による質問回答システムの音声認識部, 日本音響学会研究発表, 2.4.21. (1975年10月予定)