

# 自動通訳システム INTERTALKER における 日本語音声合成

6 P - 8

岩田和彦, 三留幸夫, 宮本 牧, 渡辺隆夫  
日本電気(株) C&C情報研究所

## 1. まえがき

自動通訳における音声出力部には、まず、任意の内容を出力できること、できるだけ明瞭で自然な音声を出力できることなどが要求される。我々は、日本語の文章を音声で読み上げるテキスト音声合成の研究に取り組んでいる。先に、人間の肉声に近い高品質な合成音声を出力することが可能な残差制御型音声合成方式を提案し、この方式を用いたテキスト音声合成システムを開発した[1]。これにより、任意の日本語テキストを明瞭な音声で読み上げることが可能である。

この他、自動通訳のような音声対話系における出力音声には、会話音声としての自然性が望まれる。音声対話系においては、発話文をシステムが生成するので、その統語構造だけでなく、意味内容も明らかになっている。また、対話における焦点を知ることが可能である。このような情報を利用することによって、発話の内容に適したイントネーションやリズム、テンポと言った韻律の制御が行えるものと期待される[2][3]。しかし、これらについては、現時点はまだ基礎検討の段階である。

そこで、我々は、先に開発したテキスト音声合成システムを、自動通訳の日本語音声出力部に適用することを試みた。本報告では、自動通訳システムにおける日本語音声合成部の概要、ならびに、自然なリズムを実現した継続時間長制御方式、高品質な合成音声が得られる残差制御型音声合成方式について述べる。

## 2. システム概要

本システムは、図1に示すように、パーソナル・コンピュータとその拡張スロットに実装された規則音声合成ボードとからなっている。音声合成ボードは、韻律情報の生成などを行う16bitマイクロ・プロセッサをCPUとし、音波形の生成を行う16bit DSP、音声データを格納するメモリなどで構成されている。機能的にはパーソナル・コンピュータ上で行われるテキスト処理、音声合成ボード上で行われる韻律制御、音声生成に大別される。

### 2.1. テキスト処理部

翻訳部から送られてくるテキストを、音素記号と、アクセントやポーズなどを表す韻律記号とからなる発音記号列に変換する。このとき、複合語や付属語の連鎖によるアクセント結合の処理、ポーズ挿入位置を決定する処理などを行う。

### 2.2. 韻律制御部

テキスト処理部で生成された発音記号列に基づいて、イントネーションやリズムのバタンを、それぞれの生成規則にしたがって決定する。

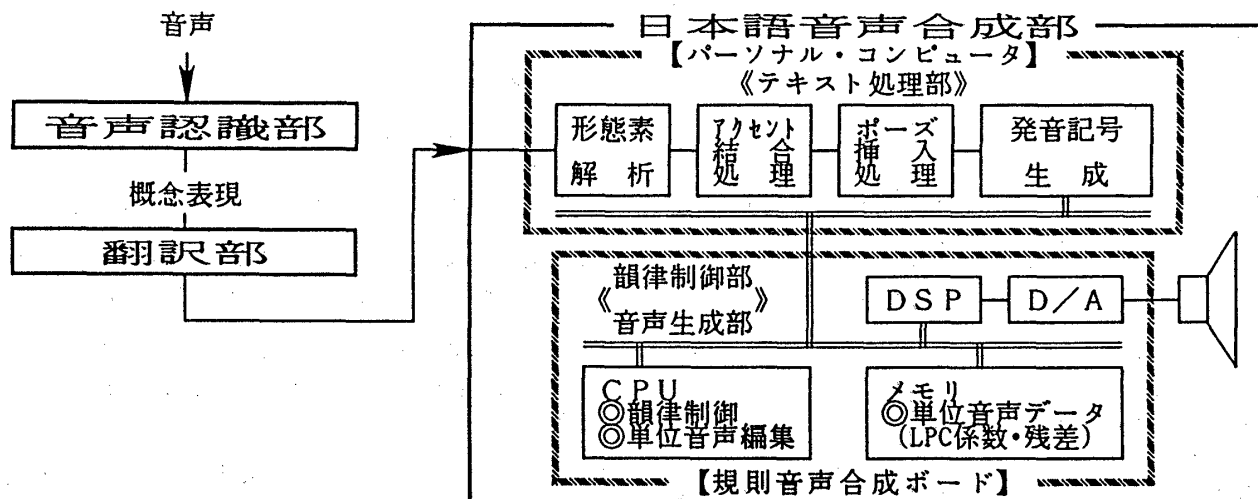


図1. 自動通訳システムにおける日本語音声合成部

### 2.3. 音声生成部

連続音声を合成するために用いる音声データとして、男女各1名のアナウンサーが発声したCV連鎖（C：子音，V：母音；50音に相当）、及びVC連鎖の音声をを用いている。男声/女声の選択は外部から行うことができる。自動通訳においては、英語の話者の性別に合わせて切り替えを行っている。音声の合成方式には、新しく提案した残差制御型音声合成方式を用いている。本方式により、明瞭性の高い合成音声が出力される。

### 3. 継続時間長制御

発話のリズムに関わる連続音声の中の各音素の継続時間長は、前後の音素の種類や文中での位置など、その音素の置かれた環境の影響を受けることが知られている[4]。そこで、各音素の継続時間長を、その音素の置かれた環境ごとに測定したテーブルを用意し、合成時には、テーブル中から環境の一致するものを検索して継続時間長を決定する方法を提案した。まず、継続時間長に対して及ぼす影響が似ていると考えられる音素をまとめたカテゴリを定義し、先行、当該、後続の各音素のカテゴリが一致する音素連鎖における当該音素の継続時間長の平均値、及びそれぞれが出現した呼気段落（ポーズで区切られる発話の単位）内モーラ数の平均値を求めた。結果の一部を表1に示す。一方、予備の実験の結果、呼気段落内モーラ数の増加にともなう継続時間長の減少は、各音素ごとに固有の傾きを持った直線で近似できることがわかった。

以上のことから、先行音素に  $P_p$ 、後続音素に  $P_f$  を持つ音素  $P_c$  の、 $M_t$  モーラの呼気段落内における継続時間長  $D_{P_p, P_c, P_f}^{M_t}$  を、次式によって表す継続時間長モデルを提案し、これにより自然なリズムが実現されることを確認した。

$$D_{P_p, P_c, P_f}^{M_t} = D_{P_p, P_c, P_f}^{M_0} + \alpha_{P_c} \cdot (M_t - M_0)$$

ここで、

$D_{P_p, P_c, P_f}^{M_0}$  :  $M_0$  モーラの呼気段落内における継続時間長 (表1)

$\alpha_{P_c}$  : 音素  $P_c$  の継続時間長の、呼気段落内モーラ数への回帰係数

### 4. 残差制御型音声合成方式

合成音声の明瞭性を向上させる方法としては、合成フィルタの励振音源にフィルタ係数の推定誤差（残差）を使用することが考えられる[5]。従来は、周期性のない音声区間では全ての残差を用い、母音など周期性のある区間では代表的な一つのピッチ区間の残差を繰り返し用いる方法が採られていた。しかし、代表ピッチ区間の残差とフィルタ係数とは対応したものでなくなるため、音質の劣化は避けられない。そこで、周期性のある区間に

表1. 継続時間長テーブル (一部)

先行	当該	後続	平均 呼気段落内 モーラ数	継続時間長 [ms]
\$	a	i	13.0	85.0
P	a	UF	12.5	89.4
US	a	UF	11.8	78.6
US	a	\$	13.8	61.5
UF	a	P	10.0	103.8
\$	US	a	11.0	38.2
u	US	a	13.4	47.7

a, i, u : 母音

US : /p, t, k/ UF : /s, ch, ts/

\$ : 文頭、文末 P : ポーズ

においても、全ての残差を用いる方式を提案した。周期性のある区間においては、ピッチ区間ごとに残差に対して零詰めや打ち切りを行うことによってピッチ周波数を変更し、イントネーションの制御を行う。

### 5. むすび

自動通訳システムの日本語音声出力部に、パーソナル・コンピュータと、規則音声合成ボードとからなる音声合成システムを適用した。

現行システムでは、出力音声の生成に概念の情報は反映されていないが、概念情報の利用は自然な音声対話系を実現する上で重要と考えられる。対話の流れや概念と、韻律などの音響的特徴量との関係を明らかにしていくことが、今後の課題である。

### 参考文献

- [1] 岩田, 三留, 友竹, 赤松, 亀谷, 小澤, 渡辺: “残差制御型音声合成方式を用いた日本語テキスト音声変換システム”, 信学技報, SP90-56 (1990-11)
- [2] 坂井, 村木, 岩田: “中間言語からの韻律情報の生成”, 音響学会講論集, 2-4-9 (1990-03)
- [3] 柴田, 山下, 溝口: “概念表現からの音声合成における対話管理”, 音響学会講論集, 3-P-18 (1991-10)
- [4] 匂坂, 東倉: “規則による音声合成のための音韻時間長制御”, 信学論, J67-A, 7, pp.629-636 (1984-07)
- [5] 佐藤: “音韻連鎖と残差波形を用いた音声合成”, 音響学会講論集, 1-2-16 (1981-10)