

## 文・音声変換技術を用いた 原稿読合せ法の有効性について

壁谷 喜義      影山 五良      石川 浩一郎

(日本電信電話公社      横須賀電気通信研究所)

### 1. はじめに

近年、オフィスオートメーションの必要性が増大するにつれ、データ宅内装置、オフコン等OA機器の機能、性能および、マンマシンインターフェースの向上が強く望まれてきている。その一手段として、音声によるオペレータガイダンス出力、ワードプロセッサで作成した文書の音声読合せ等、音声出力技術を利用した種々の補助機能を付加することにより、作業能率の改善が期待される<sup>1)</sup>。

音声を出力情報媒体とする利点として、

- 1) 手足、視線の動きに拘束されず音声情報を得ることができる。
- 2) 視覚情報等と併用することにより大きな効果を発揮する。

などが挙げられる。

例えば、操作ミスをした場合、警告をディスプレイ画面に表示する代わりに、音声メッセージで通知することにより、その時の視線の位置に拘わらず、ミスを容易に知ることができる。また、入力された文書の確認方法として、ディスプレイ画面と原稿を見比べて確認する(目視による照合)代りに、音声に変換された入力文書と原稿を読合せ照合することにより、視線の大きな移動がなくなり、文書確認の効率化を期待できる。

このうち、特に、音声による原稿読合せ機能を実現するためには、文字コード表記された任意の日本語文章を、音声に変換する文・音声変換技術が必要となる。

本稿では、まず上記、OA機器における音

声出力機能の実現をねらった、簡易でかつ効率的な文・音声変換処理の技術的ポイントを述べる。そして、次に、本技術を用いた入力文書の音声読合せ法の有効性を確認するために行った操作性評価実験の方法及び、その結果について述べる。

### 2. 文・音声変換処理

#### 2.1 前提条件

過去に、報告されている音声読合せ機能付きワードプロセッサで実現している日本語文章の音声変換方式の特徴は、以下の通りである<sup>2)</sup>。

- (1) かなはそのまま読むが、漢字は予め対応付けられた一定の読みを与える。
- (2) 単音節編集した音声で出力。
- (3) アクセントは付与せず、一定の声の高さで出力。

この方法では、処理プログラム、変換辞書等の規模はあまり大きくないが、各漢字の読みは必ずしも正しい読みとは限らない。また、アクセントが付与されないことから漢字(カ<sup>ン</sup>ジ)、幹事(カ<sup>ン</sup>ジ)の様な同じ読みで異なるアクセントを有する語の聴き分けができないなど、作成文書の読合せ用音声として十分機能しない。

そこで、筆者らは以下に示す前提条件を満足し、応用範囲の広い文・音声変換技術を検討した。

- (1) 漢字の読みは、単語辞書を用いてできるだけ正しい読みを与える。

- (2) 連続音声を合成する。
- (3) アクセントを自動的に付与する。

### 2.2 処理の全体概要

図1に文・音声変換処理の概要を示す。処理内容は、大きく漢字かな変換、アクセント付与、規則合成の3ステップに分かれる。

まず、漢字かな変換処理では、入力された漢字かな混り文に対し、単語辞書と照合して漢字列を単語に分割し、各単語の読みのほか、アクセント、品詞情報などをつける。この時、単語辞書と照合できなかった単語に対しては、漢字辞書（漢字に対する読みの情報を有する）を用いて一字単位に読みをつける。次に、アクセント付与処理では、単語のアクセント、文法情報などをもとに文節単位のアクセントを付与するとともに、ポーズの挿入や話し言葉変換（例：ワタシハ トウキョウハイック。→ワタシワ トーキョーエイック。）を行う。そして、最後に以上で得られた音韻、韻律情報をもとに規則合成処理により連続音声を合成する。

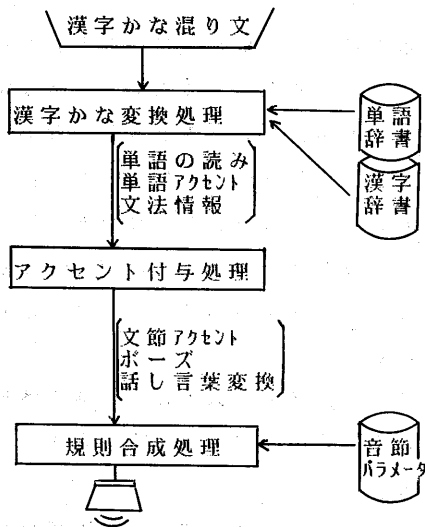


図1 文・音声変換処理の概要

以下、漢字かな変換、アクセント付与、規則合成処理の順に説明する。

### 2.3 漢字かな変換処理

表現法からみた日本語文章の性質を以下に示す。

- (1) 漢字の種類が多い。
- (2) 個々の漢字は読みを一般に複数有し、単語単位で正しい読みが決まる。
- (3) 単語毎、文節毎に分かち書きする習慣が無い。

これらの性質より、べた書きされた日本語文章中の漢字列より単語を抽出する効率的な手法が、正しい読みを与える割合を向上するためのキーポイントとなる。

いま、漢字で記述される単語約45,000語（一般語、固有名詞）の構成文字数分布を調べると、2文字以下で構成される単語が90%以上を占め、特に2文字単語だけで約80%と圧倒的に多いことが分かった。

そこで、漢字列の単語分割規則は、図2に示す通り、漢字列の前方より2文字ごとの分

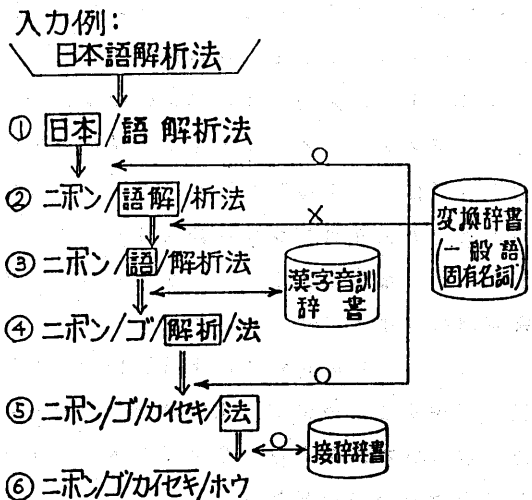


図2 漢字列の単語分割規則

割・単語辞書照合を基本とした。この場合、単語辞書と照合できなかった単語は、漢字辞書を用いて前方1文字の読みをつける(図2 ②~④)。ただし、この文字が漢字列の先頭の文字であれば、接頭語の候補として接辞辞書との照合により読みをつける。また、漢字列の最後尾に残った1文字は、接尾語の候補として接辞辞書との照合により読みをつける(図2 ⑤, ⑥)。

尚、同形異音語については、後続するかな語尾(かな語幹, 付属語, 活用語尾)の接続性判定の文法処理により一意に定める。

数字, 英字の読み方は, それぞれ, 1桁読み, スペル読みを基本とした。ただし, 数字は, 外部指定により位取り読みも可能とした。

例: 34 → サンヨン (1桁読み)  
       ↳ サンジュウヨン (位取り読み)  
    N T T → エヌディーディー

本処理方式を技術論文, 随筆など計110文(漢字含有率, 約40%)を対象として評価した結果, 97.8%の割合で正しいかな読みが与えられた。ただし, この時用いた単語辞書は約30,000語(一般語約20,000語, 固有名詞約10,000語)である。

#### 2.4 アクセント付与処理

各単語の属性情報(読み, 品詞, アクセント

形, 音節数等)をもとに, (1)複合語単位, (2)文節単位と順次アクセントを結合する<sup>3)</sup>ここでは, 音声の自然性向上への貢献度の大きい複合語アクセント規則の概要を述べる。

2単語より構成される複合語は, 構成単語の属性情報(特に後ろの単語のアクセント形, 音節数)によりアクセントが決まる<sup>4)</sup>表1にその基本規則を例と共に示す。

3単語以上より構成される複合語は, 一般に複数のアクセント句を有するが, その規則は, 音声言語学上, まだ明らかでない。そこで, 複合語アクセントの統計分析データに基づき, 構成単語数, 品詞情報を利用したアクセント句の抽出規則を検討した。以下にその規則を示す。

(1)複合語の先頭の語が固有名詞の場合, この固有名詞を分離して, 一アクセント句とする。

例: ニホン / デンシン / デンワ / コウシャ  
       ニホン\* デンシン / デンワ / コウシャ  
           " \* " = アクセント句の切れ目

(2)4単語以上より構成される複合語は, その中央で2分割する。

例: アン / オンセイ / ヘンカン / シヨリ  
       アン / オンセイ\* ヘンカン / シヨリ

(3)以上で得られた各アクセント句内で, 表1に示した2単語より構成される複合語に対するアクセント付与規則を前方より順次適

表1 アクセント付与規則

後ろの単語の属性情報		複合語のアクセント位置	例
2音節以下		前の語の最終音節	ヨリスカ・シ → ヨリスカシ
3音節以上	平板, 頭高, 尾高	後ろの語の先頭位置	シヨリ・ホウシキ → シヨリホウシキ
	いずれかのアクセント		アン・オンセイ → アンオンセイ
	上記以外のアクセント	後ろの語のアクセントを保存	タマツリ・リヨウシキ → タマツリリヨウシキ

用して一アクセント句に結合する。

例：ニホン\*デンシン/デンワ/ゴウシヤ

ニホン\*デンシンデンワ/ゴウシヤ

ニホン\*デンシンデンワゴウシヤ

ア／ン／オンセイ\*ハンカン／シヨリ

ア／ンオンセイ\*ハンカンシヨリ

アクセント付与規則を漢字かな変換処理の評価で用いた 110文に含まれる 1,450文節で評価した結果、79.2%のアクセント一致率（アナウンサー発声の音声より聴取した基準アクセントと比較）を得た。

## 2.5 規則合成処理

漢字かな変換、アクセント付与処理より得られた日本語文章の読みを表わす音韻記号系列、およびアクセント、ポーズ等韻律記号系列から連続音声を合成する。

ここで用いた規則合成処理方式は、制御および規則の簡易化がはかれるCV（子音+母音）音節を単位とする音声合成方式<sup>5)</sup>である。また、音声パラメータとしてはLSP<sup>6)</sup>を用いている。

図3に規則合成処理の概要を示す。

入力された音韻記号系列（読み情報）に従って、CV音節ファイルより順次LSPパラメータを読み出し、音韻結合規則によりLSPパラメータの結合を行なう。また、各音韻の継続時間長も規則により制御される。次に、韻律記号系列（アクセント、ポーズなど）により、ピッチ、振幅等の音源情報が規則により生成される。そして最後に、LSPパラメー

タおよび、音源情報によりLSP合成フィルタを駆動して音声を合成する。

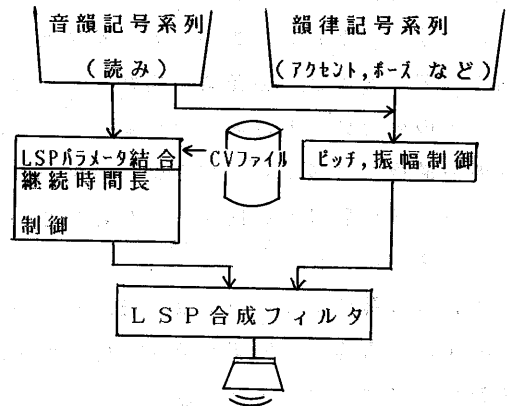


図3 規則合成処理の概要

## 3. 操作性評価実験

2章に概略を述べた文・音声変換技術の一応用としてワードプロセッサにおける、入力文書の音声読合せ法を取り上げ、その有効性を確認するため以下に示す操作性評価実験を実施した。

### 3.1 評価方法

#### 3.1.1 実験環境

音声による原稿読合せ機能を具備するワードプロセッサを模擬するため、文書作成・更新部に市販のワードプロセッサ（以降、WPと略す）を、音声出力部にテープレコーダを用いた実験システムを構成した。また、被験者の動作（テープレコーダ、WPのキー操作等）を解析するためビデオカメラを設置して録画・録音できる様にした。図4に実験システムの構成を示す。

#### 3.1.2 評価対象文章

評価に用いた計12の文章の、分野種別、

文の数、文字数などの内訳を表2に示す。このうち単行本1-①の文章内容の一部を図5に示す。

これらの文章を、実験に先立ってWPに入力しておいた。その際、各文章とも12ヶ所(1文平均1ヶ所)に入力誤り(誤字、脱字など)を意図的に設定した。

漢字の読み誤りを含む割合の差による評価の相違を調べるため、正しい読みの割合の平均をそれぞれ、(A)100%(B)98%(C)81%とした3通りの方法を実施した。

方法A, Bはともに文・音声変換技術を用いるものである。

方法Aは漢字かな変換に関し究極の目標値, 方法Bは前述の簡易な漢字かな変換処理による値である。また, 方法Cは各漢字にそれぞれ一定の読みを与える方法により得られる値を用いている。

表2に示した対象文章のうち, 方法A, B, Cの実験に用いた文章は, 各分野それぞれ①, ②, ③である。

### 3.1.3 実験手順

被験者は提示された手書き原稿の文章内容を理解するため, まず原稿を黙読する。次に, 原稿を見ながら, WPに表示された同一文章(入力誤りを含む)の内容をテープレコーダの音声(4.8Kb/s LSP-CV合成音声)で確認する。そして, 入力誤りを発見した時点で, 直ちにWPディスプレイ画面上で文章を修正する。

尚, テープレコーダで使用できるキーは, ①PLAY(音声出力開始), ②PAUSE(一時停止), ③頭出し(文頭復帰後出力) ④REWIND(任意位置に復帰後出力)である。

実験は研究室員3名を対象に1人1回3通りの方法(3文章)を, 計4回実施した。文章の提示順序は分野種別, 実験方法に関して,

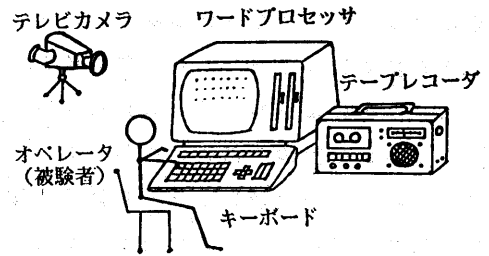


図4 操作性評価実験システム

表2. 評価対象文章の内訳

	分野種別	文の数(数)	平均文字数/文	漢字含有率* (%)	
1	単行本	①	15	41	36.2
		②	11	56	29.9
		③	18	34	28.7
2	広報文	①	11	57	31.9
		②	13	47	35.7
		③	12	52	34.3
3	技術論文	①	9	69	38.3
		②	11	56	38.0
		③	10	63	45.0
4	新聞記事	①	6	104	25.6
		②	12	52	40.1
		③	6	104	41.9

$$* \text{漢字含有率} = \frac{\text{漢字文字数}}{\text{全文字数} - \text{記号文字数}} \times 100$$

電話は「耳型」の通信メディアである。相手の耳に、音声をそのままに伝えるだけでなく、雰囲気も伝えることができる。人間の意志疎通に便利な通信手段である。手紙になると、文字が読めるかどうか、読めても言葉の意味がわからないこともある。この「耳型」の特性が、電話社会を招来した原因の一つといえる。

電話は話している間、相手を拘束している。人間は聖徳太子のように一人で十人の訴えを同時に聴くというたとえ話は例外として、音声による会話のときに同時に複数のことは聞けない。相手と話している間は、相手を拘束しており、それが電話の基本的な効用であるリアルタイムの情報授受に発展する。情報をただちに相手に伝える。ただちに答えを求め、

図5 評価対象の文章の一部  
(「テレコム革命」より)

できるだけランダムになる様配慮した。

尚、従来多く実施されている原稿の照合法である、目視による照合法との比較を行うため、第1回目の実験の前に、原稿とWPディスプレイ画面との比較による文書修正実験を1文章について行った。この時用いた文章の長さ、および設定した入力誤り状況については、表2に示した音声読合せ評価用文章と同等にした。

原稿の音声読合せ、および目視による照合法実験により、以下に示すデータを主に収集した。

- (1) 修正に要した時間。
- (2) 修正率。
- (3) キー操作回数(主にテーブルコーダ)
- (4) 習熟度。

### 3.2 評価結果と考察

#### 3.2.1 修正時間と修正率

音声読合せ法による修正時間と修正率の実験回数による推移を正しい読みの割合の異なる文章ごとの実験(方法A, B, C)について図6に示す。さらに、表3に、4回の実験の平均修正時間、平均修正率を示す。ここで、修正時間とは、音声読合せ開始から終了までの合計時間のうちテーブルコーダの巻戻し・頭出しに要した時間を除いたものである。また修正率とは、12ヶ所の全入力誤りのうち、その誤りを発見できWP上で修正した箇所の割合である。

この結果、目視による照合法に比べ、音声による原稿の修正確認では、方法A, B, Cすべて修正時間の短縮(平均20%)をはかることができた。また、方法A, Bでは目視による照合法に比べ、修正率においても改善

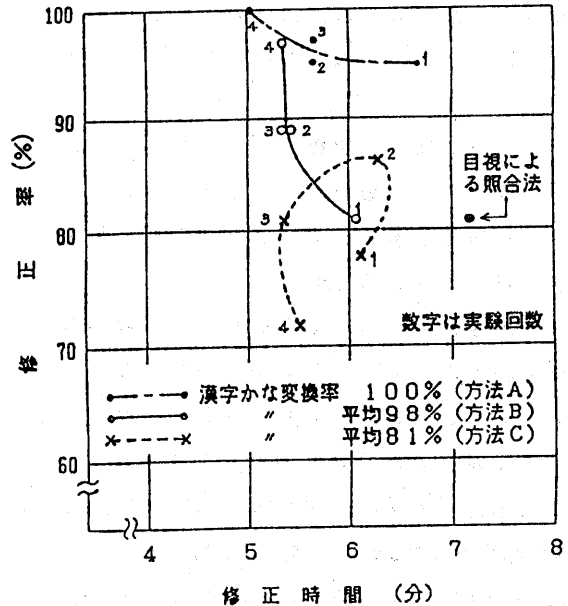


図6. 修正時間と修正率

表3. 修正時間・修正率の平均

方法	平均修正時間	平均修正率(%)
A	5分43秒	96.5
B	5分33秒	88.9
C	5分49秒	79.3
目視による照合法	7分9秒	80.6

された。しかし、方法Cでは、修正率の改善は得られず、方法A, Bに比べあまり有効な方法とは言えない。また図6より、方法A, Bでは、音声読合せ作業の回数を重ねるに従い修正時間の短縮、修正率の向上が期待できる。

#### 3.2.2 一時停止キー押下回数

音声出力中、テーブルコーダの PAUSE (一時停止) キーを押下する要因は、

- 1) 入力誤りを発見した時。
- 2) 漢字かな変換のミスによる読み誤りを生じた時。

3) その他(聞き逃し、不明瞭な音節等)である。そのうち、1)は、文書修正のための有効なPAUSEである。そこで、いま「有効なPAUSEキーの押下比率」を(1)式の通り定義する。

$$\frac{\text{有効なPAUSEキーの押下比率} = \text{修正箇所でのPAUSEキー押下回数} \times 100\%}{\text{PAUSEキー押下の全回数}} \dots (1)$$

図7に、方法A、B、Cにおける有効なPAUSEキーの押下比率およびPAUSEキー押下の全回数を1文章平均して示す。方法A、Bは方法Cに比べ、PAUSEキー押下回数が少く、かつ有効なPAUSEキーの押下比率が大きいことがわかり、疲労の軽減、操作性向上を期待できる。

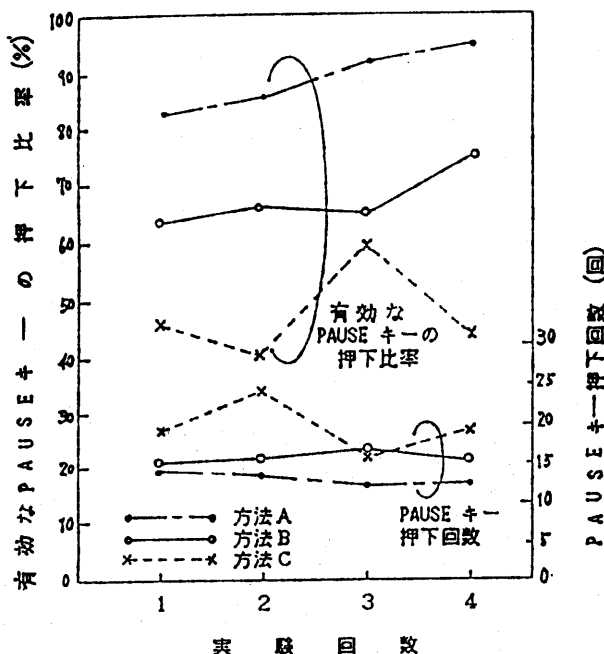


図7. PAUSEキー押下回数の比較

#### 4. むすび

本稿では、文・音声変換技術を用いた音声による原稿読合せ法の有効性について論じ、以下の結果を得た。

- (1) 目視による照合法に比べ、約20%の修正時間の短縮を図ることができた。
- (2) 修正率においても目視による照合法に比べ、8%以上の改善を図ることができた。
- (3) 文・音声変換を用いることにより、各漢字に一定の読みを与える音声読合せ法に比べ、音声一時停止キーの押下回数が少なくなる。

以上の結果、文・音声変換技術を用いた原稿の読合せにより、文書確認・修正の能率化、および疲労の軽減が期待できる。

今後は、以上の評価結果を踏まえ、性能の向上を一層はかり、音声読合せ機能を有する装置の設計に反映していく。

終りに、日頃御指導頂く、松田宅内部長、山崎統括役、石井音声入出力方式研究室長に深謝します。また、操作性評価実験の被験者として御協力頂いた室員の皆様に重ねて感謝致します。

【参考文献】

- 1) 上田, 藤方, 江尻: 「マイクロコンピュータによる素片波形接続形音声合成方式とその応用」, 信学論(D), No.11, pp931-938, (昭55)
- 2) 中島: 「日本語ワードプロセッサBW-20における音声読合せ機能」, 昭57年度信学会総合全大予稿, 5-389
- 3) 壁谷, 石川: 「文音声変換におけるアクセント付与規則の検討」, 情処自然言語処理研究会 ('82.5)
- 4) 日本放送協会編: 「日本語アクセント事典」, (昭43)
- 5) 東倉, 匂坂: 「CV音節を単位とする音声合成」, 音講論, 3-4-3, 1980年5月
- 6) 板倉, 管村: 「LSP音声合成器の原理と構成」, 音響学会音声研資, S79-46, 1979年11月
- 7) 北原安定: 「テレコム革命」, 徳間書店(昭58)