

自動プレゼンテーションシステム(3)

1 Q-3

～プレゼンテーションの実行～

山口 浩司 長谷部 浩一 小野口 一則 金澤 博史 恒川 尚

(株)東芝 総合研究所

1. はじめに

我々は文書を理解しその内容をプレゼンテーションするシステム(APS)の試作を行った[1][2]。

本稿では、文書入力処理部[3]で出力・格納された文書データに基づいて実際にプレゼンテーションを行う際の各処理部での機能について述べる。

2. プレゼンテーションの進行

プレゼンテーションでは、入力文書の内容に沿って、文章については箇条書きに、図表はより見易く変換して聴衆に呈示する。同時にその説明文をナレーションとして発声しながら進行する。このとき呈示資料は、段落あるいはフレーズ単位にナレーションと同期させて強調表示するなどきめ細かく変化する。表データは必要に応じて各種のグラフに変換される。

3. 表データのグラフ化

3.1 データ構造

今回扱った表は、二次元配列の構造を有するものである。表データは、表枠内の左端/上端にある縦/横の行/列見出し部と数値データ部、ならびに表枠外の表見出し部により構成される(図1)。これらは、表認識処理部によって識別され、図2に示すデータ構造で格納されている。表データ解析部が、表枠内各セル内の文字列の形態解析を行って、行/列見出し部および数値データ部の判別と各セルの文字列から項目名、数値単位、数値データを抽出する。

行/列見出し部は、それぞれ行/列の項目名や数値データの単位を示す文字列を含む。図2では、「旅行者数」や「旅行収入」などのほかに年次を表す「1985年」などの文字列が項目名として識別されている。「百万人」、「十億ドル」などは数値単位として検出されている。これらは、箇条書き変換部の文章解析手法と同様にして、単語辞書との照合により抽出されるものである。

数値データ部は、数字列の二次元配列となっている。表データ解析部が、各セル内の数値に対して、各セル内に記述された数値単位のほか行/列見出しに含まれる数値単位のうち適切なものを対応させる。

表見出し部には、通常、表番号や表タイトル、出典に関する記述が含まれており、これらは本文中の表参照箇所と対応させることができる。表データ解析部では、箇条書き変換部での文章解析結果と照合をとることにより、文章中の表参照箇所と表データの対応付けも行っている。

3.2 グラフ化

表データとして入力された数値データを、プレゼンテーション時には、グラフ化して呈示する。今回は、グラ

表科1 世界の国際観光量					
行見出し部		1985年	1986年	1987年(推計)	列見出し部
表見出し部	旅行者数(百万人)	333.8	341.4	355.0	数値データ部
	旅行収入(十億ドル)	108.6	129.2	150.0	

(注) WTO資料による

図1 表データの構成

表見出し部				
表見出し	"世界の国際観光量"			
表番号	"表科1"			
注釈	"(注)WTO資料による"			
行数	2	列数	3	

行見出し部				
行	見出し	属性	単位	文字列
1	"旅行者数"	人数	百万人	"旅行者数(百万人)"
2	"旅行収入"	金額	十億ドル	"旅行収入(十億ドル)"

数値データ部				
(行列)	値	属性	単位	文字列
(1,1)	333.8	数値		"333.8"
(1,2)	341.4	数値		"341.4"
(1,3)	355.0	数値		"355.0"
(2,1)	108.6	数値		"108.6"
(2,2)	129.2	数値		"129.2"
(2,3)	150.0	数値		"150.0"

列見出し部				
列	見出し	属性	単位	文字列
1	"1985年"	年次	なし	"1985年"
2	"1986年"	年次	なし	"1986年"
3	"1987年(推計)"	年次	なし	"1987年(推計)"

図2 表のデータ構造

フのタイプとして、「棒グラフ」、「帯グラフ」、「円グラフ」の3種類を用意した。どのタイプのグラフを使用するかは、前述の表データ解析部の解析結果を参照し、グラフ変換規則により決定される。

たとえば、図2の解析結果によれば、各列見出しは年次を表しており、また、各行見出しは、それぞれ人数、金額という数量を表している。よって、この表は人数と金額の「推移」を示す表であると解釈される。さらに、このような推移を表すグラフとして適切なのは、「棒グラフ」であることを規則により推定する。図3に示したのは、「旅行収入」の推移を表す棒グラフである。

グラフ変換規則には、各表データの行/列見出しの組合せによりその表の意味する事項を判断し、「構成比率」を表した表データには「円グラフ」や「帯グラフ」を、「構成比率」の「推移」であれば「棒グラフ」を対応させるという規則が含まれている。

また、質問応答時に聴衆の好みに応じて動的に多種のグラフへの変換が可能である。複数の帯で表された「旅行者数の行先別比率の推移」の「帯グラフ」を、複数の「円グラフ」で表現することは可能である。上記3種類のグラフで表現できない表データについては、入力画像と同じ表形式で整形出力される。

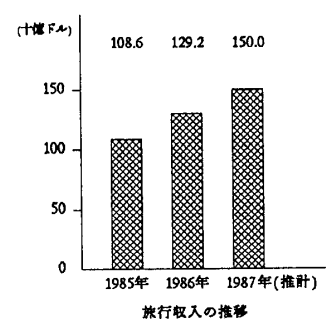


図3 グラフ化の例

4. ナレーション

簡条書き変換と同時に、入力文章からナレーション用文章が生成される。これを呈示資料画面と同期させながら合成出力する。このようなナレーションの合成に使用する音声合成の方式は、発声の自由度という点から規則合成が望ましい。しかし、プレゼンテーションのように長時間にわたり複雑な内容を伝達することが目的の場合には、聞き取り易いことも必須の条件である。残念ながら現状の規則合成技術はそのレベルに達していない。実用イメージを試作するという方針から、原稿の内容を限定して、より滑らかな発声可能な単語単位の録音再生方式を採用した。記憶容量は1800語で100Mバイトである。

5. 聴衆モニタ

システムを簡素化するためには1台のカメラで聴衆をモニタすることが望ましい。着席判定や挙手判定の処理を1台のカメラから入力される画像を用いて行うためには、着席している人の顔や挙手された腕が同時に入力できる位置にカメラを配置する必要がある。このため、客席を斜め下方に見下ろす位置にカメラを設置し、このカメラから入力される画像(図4)を用いて観客席をモニタした。

5.1 着席判定

入力画像では、前列に着席した人が後列の座席を隠すため、空席時の画像と入力画像との間の座席部分の単純な差分を求めるだけでは、座席が空いているか否かの判定は難しい。このため、ITVカメラから入力されるRGBの色情報から画像中の肌色領域(顔の部分に対応)及び黒色領域(頭髪の部分に対応)を抽出し、この組合せにより着席している人の頭部を求め、座席に人が座っているか否かを判定した。

各座席は固定されていると仮定し、まず、座席に人が座っている際に、頭の部分が存在する領域(以下、頭部存在領域と呼ぶ)を設定する(図4の矩形領域)。次いで、この領域内の各画素(x,y)に対し、以下の式を用いてR,G,B各画面の揮度 $IR(x,y)$, $IG(x,y)$, $IB(x,y)$ の比 $RR(x,y)$, $RG(x,y)$, $RB(x,y)$ を求める。

$$RR(x,y) = IR(x,y) / (IR(x,y) + IG(x,y) + IB(x,y))$$

$$RG(x,y) = IG(x,y) / (IR(x,y) + IG(x,y) + IB(x,y))$$

$$RB(x,y) = IB(x,y) / (IR(x,y) + IG(x,y) + IB(x,y))$$

そして、プレゼンテーション時と同一の照明環境で予め設定しておいた肌色に対応するRR, RG, RB, IRの閾値 $RRmin$, $RRmax$, $RGmin$, $RGmax$, $RBmin$, $RBmax$, $IRmin$, $IRmax$ を用い、

$$RRmin < RR(x,y) < RRmax$$

$$RGmin < RG(x,y) < RGmax$$

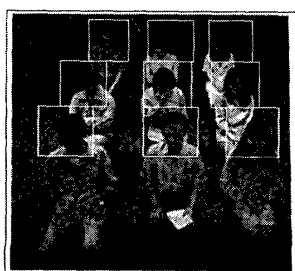


図4 着席判定

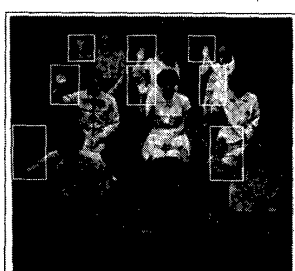


図5 挙手判定

$$RBmin < RB(x,y) < RBmax$$

$$IRmin < IR(x,y) < IRmax$$

なる条件式を満足する画素(x,y)を肌色と判定する。黒色に対しても対応する閾値を設定し、上記の肌色抽出処理と同じ手法を用いて各画素が黒色か否かを判定する。

上記手法を用いて頭部存在領域内の肌色及び黒色領域の画像を生成し、ラベリング処理後、ヒストグラムをとることにより頭部存在領域内における最大面積の肌色領域H及び黒色領域Kを抽出する。次いで、HとKの外接長方形を求め、各長方形の重心間の位置関係を求める。この位置関係及びH,Kの面積が、各頭部抽出領域において予め設定しておいた値に近い場合に、領域Hを顔、領域Kを頭髪とし、座席に人が座っていると判定する。

5.2 挙手判定

挙手する以前の画像 Pp と挙手している際の画像 Pa との間の差画像 Ps を用いて挙げられている手を抽出する。挙げられている手の部分の領域は Pp と Pa の間で輝度的に変化しているため、適当な閾値で差画像 Ps を2値化すれば挙手された手の部分の領域が抽出できる。ただし、聴衆は常に僅かながら体を動かしているため、手の部分以外の領域も Ps 中には多数現れ、これらを除去する必要がある。このため、手の平に対応する肌色部分の領域を画像 Pa において抽出し、この領域に含まれ、かつ Ps において抽出された揮度変化領域にも含まれる領域を挙手されている手として抽出した。肌色部分の抽出は空席判定と同じ手法を用いた。図5に実験結果を示す。抽出された腕が矩形で囲まれている。聴衆は半袖や長袖等の様々服装をしているが、挙手前の画像との変化領域において手の平に対応する部分を求めることにより、正しい挙手判定が行われている。

6. 質問応答

プレゼンテーションの際に、聴衆からの質問を受け付けることも本システムの必須機能である。この点で、不特定話者、連続発声、騒音下、多数単語など音声認識技術に対する要求は多い。同時に、高度な文章理解技術も必要となる。しかし、現状の技術ではこれらをすべて満たすことはできない。ここでは、質問の内容の自由度に重点をおいて、連続発声、多数単語の実現を図った。質問者は特定して、連続発声された単語を音韻を単位として認識する音韻遷移照合方式[4]を採用した。

7. まとめ

本システムでは、聴衆の動作をモニタして着席判定と挙手判定を行い、その結果をダイナミックにプレゼンテーションの進行に反映させた。また、音声によるナレーションを呈示資料の表示と同期させることで視覚および聴覚によるユーザインタフェースを実現することができた。

参考文献

- [1] 恒川他: マルチメディア情報の理解と相互インタフェース, 情報処理学会「マルチメディア情報と分散協調」シンポジウム論文集, pp.67-76 (1989).
- [2] 長谷部他: 自動プレゼンテーションシステム(1), 本大会予稿 (1990).
- [3] 堤竹, 鈴木, 堀, 恒川: 自動プレゼンテーションシステム(2), 本大会予稿 (1990).
- [4] 坪井, 竹林: 静的・動的特徴ベクトルを用いた連続音声認識方式の検討, 音講論, 3-1-6 (1989).