

マルチメディア通信・情報処理の研究開発

木戸出 正継 坪井 宏之 駄竹 健志

(株)東芝 関西研究所

658 神戸市東灘区本山南町8-6-26

要約: 1992年3月神戸市本山に、マルチメディア情報に関する通信ネットワーク技術とパターン理解・知的信号処理などの情報処理技術の研究開発拠点として、関西研究所を開設した。研究環境はインテリジェントビル内にTV会議などの映像通信やLANなどのコンピュータ通信のネットワークと画像専用のコンピュータや一人一台のワークステーション群で構成されている。これらの先端設備をフルに活用し、高度情報社会の実現に必要なマルチメディア通信・情報処理要素技術の先行研究開発を進めると共に、この分野における国際的にも高度なレベルにある大学や研究機関との交流を深め、地域社会との融合を図っている。

具体的研究として、コンピュータによる自然な文章読み上げ・対話を目指す音声合成技術と通信帯域が極端に制限された場合にも有効な超低レート動画像圧縮技術の現状を紹介する。

Advanced Research Laboratory for Multi-media Communications and
Information Processing Technologies in the 21st Century

Masatsugu Kidode, Hiroyuki Tsuboi and Kenshi Dachiku

TOSHIBA Kansai Research Laboratory

8-6-26, Motoyama-minani-machi, Higashinada-ku, Kobe, 658

Abstract: A new research and development base, Kansai Research Laboratory, was established for multi-media communications and information processing technologies of the 21st century in March 1992. With the aids of advanced computing and networking facilities, we are studying several key technologies for next generation's communication networks, multi-media information processings at man-machine interfaces and multi-dimensional signal processings; for example, new speech synthesis and image compression techniques. And also we are developing relationships with several research institutes and universities in Kansai area.

1. はじめに

21世紀の高度情報社会の実現を目指して、マルチメディア情報に関する、大量・超高速通信ネットワーク、パターン理解及び合成によるヒューマンインタフェース、知的信号処理及び評価などの要素技術の先行研究開発に取り組む拠点として、1992年3月神戸市本山に関西研究所を設立した。同時に、関西システムセンターも開設し、主に関西地区におけるコンピュータ応用システムや情報通信システムなどのソフトウェア開発を行い、研究所と共に関西地区における技術開発の推進拠点として活動している。

研究所は地上14階建てのインテリジェントビル内にあり、TV会議などの映像通信やLANなどのコンピュータ通信のネットワーク設備及び画像専用コンピュータや1人1台のワークステーションを備えている。これらの設備をフルに利用し、前述のような通信・情報処理の先行研究開発を進めると共に、この分野において国際的にも高度なレベルにある大学や研究機関との交流を深め、地域社会との融合を図っている。

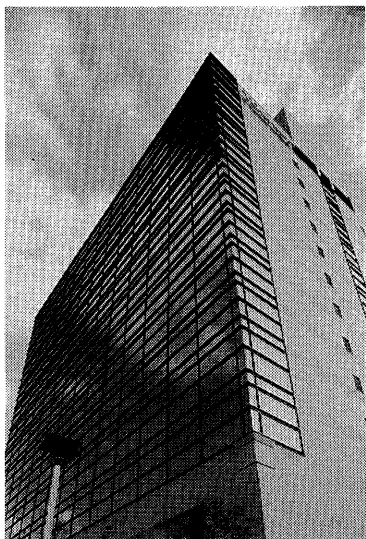


写真1 建物外観

2. 研究内容

図2.1に示すようなマルチメディア情報に関する通信・情報・画像システムの要素技術を研究開発している。

2.1 次世代通信ネットワークシステムのキーテクノロジーの研究開発

従来の電話音声データやコンピュータデータの情報伝送のみならず、HDTV画像などの大量情報までも伝送できる広帯域サービス総合デジタル網(B-ISDN)の基盤技術の一つであるATM(Asynchronous Transfer Mode-非同期転送モード)に関するトラフィック解析や網管理・制御方式や無線を利用した通信ネットワーク利用技術の研究を行っている。企業内通信網を中心に、これからのネットワーク・交換機・端末機器の要素技術の研究を進めている。

2.2 マルチメディアインタフェースの要素技術開発

音声言語理解と3次元視覚情報処理を中心に進め、ヒューマンインタフェース(HI)技術の高

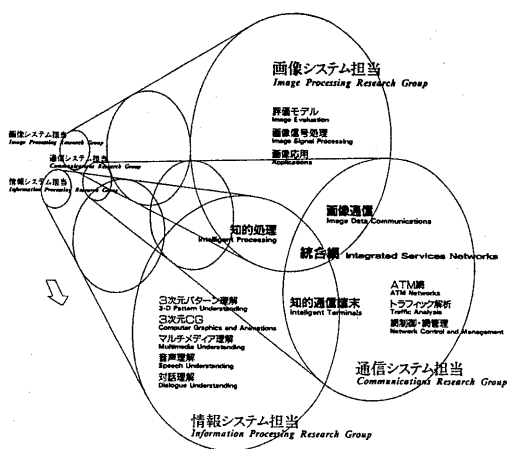


図2.1 研究分野

度化を図っている。

音声言語理解では、対話内容の言語理解と音声応答技術の研究を進め、音声認識技術と合わせて、自然でかつフレンドリーな対話システムの開発をしている。また、3次元視覚情報処理では室内環境の画像情報から3次元形状や表面の情報を獲得し、対象をモデル化する技術と、それらの情報を自由に操作できるCG技術の研究を行っている。そして、言語・音声・画像を統合したマルチメディアHIで、人と機械が優しく、楽しくかつ賢く、付き合えることを目指している。

2.3 画像情報の知的信号処理技術の研究開発

従来の圧縮処理技術を超えた知的信号処理技術を研究し、画像情報を効率良く伝送したり蓄積できるシステムを目指している。また、処理されたり合成された各種の画像の評価を行う視覚モデルの研究を行い、信号処理技術と一体化することにより、新しい信号処理の研究分野を切り開いている。これら画像処理技術の応用を医用診断、セキュリティ制御などの分野に展開している。

3 マルチメディア研究開発の具体例

3.1 音声規則合成

音声合成は利用者に分かりやすく情報を伝達する手段の一つである。特に、音声規則合成は、任意のテキストを音声に変換できる。文・音声変換による読み上げや音声による対話など幅広い分野で音声規則合成を応用するためには、合成音声の音質、自然性の向上が必要である。合成音声の品

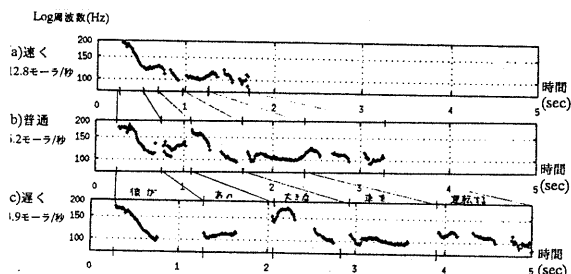


図. 3.1 発声速度の異なる発声の基本周波数の変化の例 (彼はこの車を運転する)

質には、各音の明瞭さと韻律的特徴（音の高さ、強さ、ストレス、速さなど）の自然さの2つの要素が関連する。特に、韻律の制御は、伝達する音声の内容を正しく、分かりやすく伝えるために重要な技術である。

現在、音声規則合成技術の高度化を目的として、アクセントやイントネーションの制御方式を検討している^[1]。アクセントやイントネーションに限らず韻律的特徴に関わる制御情報は、合成する文の内容に関連が深く、文の構造・意味情報を得ることが必要となる。また、構文・意味処理の結果だけで、直接韻律の制御情報に変換されるものではなく、人間の発声特有の特徴を考慮しなければ自然な音声を合成することは出来ない。例えば、図3.1のように発声速度を変えた場合、アクセント・イントネーションの変化は異なる。そこで、文の構文・意味構造と韻律の関係を検討し、構文解析の結果を音声規則合成に利用する枠組みと規則を構築する必要がある。ここでは、文の構文・意味性質と韻律の関係を整理するために、文章発声における休止位置と文の構文・意味構造の関連について分析検討した結果について述べる。

(1) 韻律と構文・意味構造の関係

日本語を含む複数の言語において、右枝分かれ構造が韻律上、有標(有意な標識)として知られている。ここで、枝分かれ構造とは句構造標識として用いられるものであり、文を構成する各要素が文中においてどの文法範疇に属するかやどのような支配関係にあるかを示す。

枝分かれ構造と韻律規則の関係は、いずれも複数の要素(語や句)を一つの単位とする働きを持つことである。文の初めから線索的に、左から右へと発話が行われている過程にあつて、意味的なまとまりを持つ構造の単位で韻律上の融合(呼気段落)が生じる。他方、意味的なまとまりを持たない右枝分かれ構造では韻律の融合が起こらない。ここでは、これを枝分かれ制約と呼ぶ。図3.2は、「韻律単位の融合を促進する」と韻律単位が融合に関わる」の2つの文の構文・意味構造(実線部分)と呼気段落(破線部分)を示している。

しかし、韻律の融合が枝分かれ制約を逸脱する場合があり、枝分かれ制約が文全体の意味に関係して

どのように働くかについては明確でない部分が多い。例えば、「私はここで出来るだけ簡素な暮らしをした」では、図3.3で示すように枝分かれ制約以外の要因が呼気段落の形成に関与している。そこで、文中の呼気段落の形成と構文・意味構造の関係を規則化する目的で、40文の文発声の休止位置を調査し、枝分かれ制約がどのように呼気段落形成に関わっているかを分析した。

(2)分析と規則化

文発声の聞き取りによって明らかにわかる休止を文上でマーキングした。その位置が文の構文・意味構造中でのどのような位置関係にあるのかを調査し、境界部分の生じた位置が文全体の構文・意味構造中でどのように機能しているかについて分析した。

- (a) 左枝分かれ構造中(構文・意味構造として無標)に休止が置かれる(韻律上有標)場合が35個
- (b) 右枝分かれ構造中(構文・意味構造として有標)

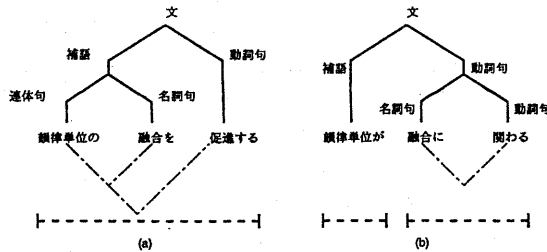


図3.2 構文構造と韻律構造の対応の例 ((a)「韻律単位の融合を促進する」、(b)「韻律単位が融合に関わる」)

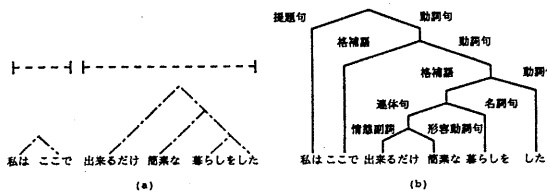


図3.3 構文構造と韻律構造の対応の例 (「私はここで出来るだけ簡素な暮らしをした」) (a) 韻律構造韻、(b)構文構造)

に休止が置かれ(韻律上無標)場合が187個

(a), (b) は構文・意味構造と韻律上とで有標位置が違っている場合を示している。40文中、構文・意味構造上で有標となる位置は119個(総有標数476個)である。

この結果を基に融合した句のモーラ数、融合した句に後続する句の構文・意味的な定性性質、融合した結果、形成された呼気段落のモーラ数ならびに句の性質を調べた結果以下のことが分った。

- ・枝分かれ規則が常に働いているのではなく、句の形態規模が枝分かれ規則を適用するかどうかに影響している。
- ・ヲ格補語は原則的に枝分れ制約が働かず、後続する句に融合する
- ・従属句は原則的に枝分れ制約が働いて、後続する句には融合しない
- ・枝分かれ制約が働かないため、ある句が融合する場合に、その句に後続する句が連体句、副詞句、従属句、提題句である場合には、融合が生じない。結果として枝分かれ制約が働いた場合と同じ結果が生じる
- ・形態規模の小さい提題句は後続する一般補語と韻律上融合する傾向にある
- ・形態規模の小さい必須補語は後続する一般補語と韻律上融合する傾向にある

本実験で得られた結果は、文の構文・意味分析結果を韻律制御に利用するために、韻律規則として音声規則合成システムに組み込んでいる。

3.2 画像符号化

動画像情報は非常に大きな情報量を有するので、画像情報を圧縮する画像符号化技術は、マルチメディアシステムにとって必須の技術である。画像符号化の標準化も盛んに行われており、テレビ会議・テレビ電話用としてH.261(64k~2Mbps)、CD-ROMなどの蓄積用としてMPEG1(1M~4Mbps)、衛星放送・ケーブルテレビなどの高画質用としてMPEG2(4Mbps以上)が標準化されている。しかし、伝送帯域や蓄積容量に制限のある場合、例えば無線による動画像伝送やメモリカードへの動画像の蓄積など、より圧縮率の高い画像符号化方式の要求が高まっている。このような要求を受けて、93

年にMPEG4として動画像を4.8k~64kbpsに圧縮する画像符号化の標準化作業が開始された。

64kbps以下の超低ビットレートで動画像を符号化するためには、動き補償と波形符号化の効率改善が必要である。われわれは、動領域全体の動き補償と三角パッチによる局所的な動き補償を組み合わせた二段階動き補償^[2]により動き補償の効率を改善した。また、波形符号化の効率改善としてDCT符号化方式より効率が高いサブバンド外挿内挿予測符号化方式^[3]を開発した。

(1)符号化アルゴリズム

図3.4に二段階動き補償外挿内挿予測符号化方式のブロック図を示す。

まず、動領域検出部で入力画像のフレーム間差分画像を用いて動領域を検出する^[4]。動領域の検出は、動物体のエッジを画面の上下左右から探索し、その内側を動領域としている。ノイズによる誤検出を防ぐために、短冊状の帯単位でエッジを探索したり、適応しきい値処理を行ったり、メディアンフィルタによる検出エッジの修正を行ったりしている。また、前回の復号画像との差の大きい領域も動領域に含めている。

次に、動領域輪郭符号化部で動領域の輪郭を符号化する。ここでは、 8×8 画素のブロックに動領域が半分以上含まれているブロックを1、それ以外を0とし、その0/1パターンをFAXで用いら

れているMR(Modified READ)符号化する。そして、動領域ブロックの各辺の中点をスプライン補間により滑らかにつないでいる。

動領域全体動き補償部では、アフィン変換を用いて、平行移動だけでなく、回転、拡大、縮小も含めて動き補償を行う。その後、動領域に三角パッチを当てて、誤差が最小になるように三角形の頂点をずらし、局所的な動き補償を行う。

次に、入力画像と動き補償画像をサブバンドに分割する。適応予測では、元の画像で 16×16 のブロックに対応する各サブバンドの入力サブバンド画像に対する動き補償サブバンド画像、前フレームサブバンド画像、背景サブバンド画像、定数0との誤差電力をそれぞれ計算し、最小のものを予測画像として選択する。背景メモリの更新は、定数0が選ばれたとき、すなわちフレーム内符号化されたときに行う。

適応予測誤差は、更に外挿内挿予測符号化される。外挿内挿予測符号化は、まず、ブロック右下の点をブロックに隣接する既に符号化された画素を用いて外挿予測し、予測誤差を量子化する。次に、中間の画素をブロック右下とブロックに隣接する画素を用いて内挿予測し、予測誤差を量子化する。同様にして間の点を埋めていく。予測係数は、各サブバンド毎に変えている。また、フレーム内予測の場合とそれ以外の場合とで予測係数を

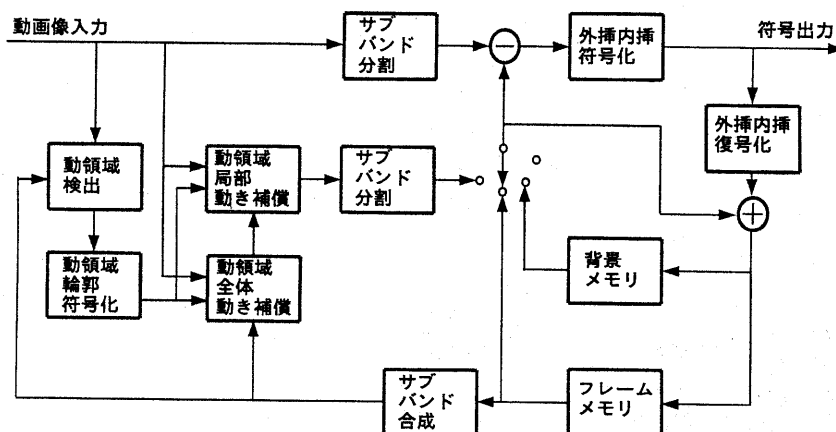


図3.4 二段階動き補償外挿内挿予測符号化方式のブロック図

変えている。

(2)実験結果

画像ClaireとMiss AmericaのQCIF画像を用いて、64kbps以上の国際標準方式H.261との比較実験を行った。量子化ステップサイズは固定で、平均8kbpsになるように調整した。図3.5にClaireにおける比較結果を示す。ビットレートはほぼ同じであるが、SNRは提案方式のほうが1~5 dB高い性能が得られている。また、H.261方式は、8kbpsまで落とすとブロックひずみやぼけがかなり目立つが、提案方式は、ブロックひずみやぼけが大幅に改善されている。

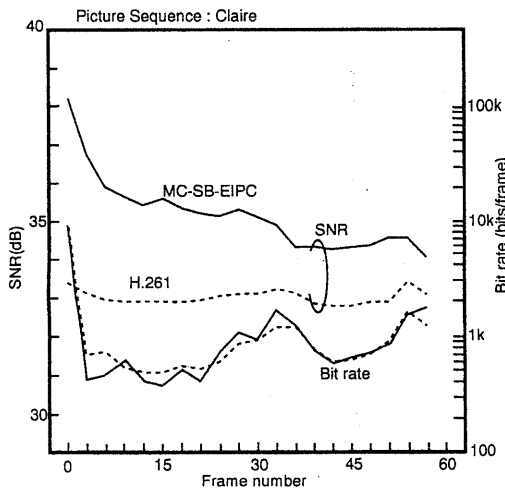


図3.5 TV会議電話規格H.261と提案方式の比較 (テスト画像はClaire)



写真2 執務風景

4. おわりに

21世紀の高度情報社会においては、人間の知覚・感覚に訴える多種の情報に対する通信・処理・蓄積などができるシステムを考える必要がある。これら要素技術の研究開発に挑戦しながら、新しい通信・情報処理研究の芽を育てる環境の構築をし、フレッシュでアグレッシブで、インターナショナルな研究所を目指して、一步一步着実に進めている。

[参考文献]

- [1]坪井宏之, 山下泰樹, 新居孝章: 「発声速度を考慮した F0 モデルパラメータの制御」 音響学会講演論文集, 1-8-21, 1993.
- [2]山口, 駄竹: "アフィン変換を用いた一般化動き補償に関する一検討", PCSJ '93, 1-7, 1993.
- [3]駄竹: "画像のサブバンド外挿内挿予測符号化", テレビジョン学会技術報告 Vol.17, No.57, pp.19-24, Oct.1993.
- [4] 駄竹, 面川: "動領域抽出に関する一検討", テレビジョン学会年次大会24-3, 1993.