

解説 Review

ユビキタスなマルチメディア・インタフェース

～ユビキタス・コンピューティングの為の最新コンテンツハンドリング技術～

中 瀧 信 弥・篠 原 章 夫*・日 高 浩 太**・菊 地 由 実**

Recent multimedia interface technologies for ubiquitous environment

Shinya NAKAJIMA, Akio SHINOHARA*, Kota HIDAHA**, Yumi KIKUCHI**

Abstract: ‘Ubiquitous Computing’ and ‘Mobile Computing’ are getting more popular. Mobile phones, for instance, enable us to listen to music, browse web pages, and watch TV. Many international airports provide ‘hot-spots’ and in major hotels, you may be able to plug your lap-top PC into the Internet. This paper defines ‘ubiquitous computing’ as the IT systems that are usable everywhere, whenever, and by whomever. The main points of this paper are the human interface technologies that can adapt to a wide variety of people, places, and times. This paper introduces three technologies: the automatic skimming interface ‘ChocoPara’, the new interactive public media ‘Mirai-Tube’, and the intelligent interface system ‘o Robo’ which can create TV-like presentations from web pages. Evaluation experiments confirm that all these technologies are effective and useful. Moreover, each methodology has extensive headroom and can be easily extended to more general ubiquitous services.

Keywords: Ubiquitous, multimedia, skimming, interactive, public place, TV, robot, web page.

1. はじめに

近年、ユビキタス・コンピューティング (Ubiquitous Computing, 以降はユビキタスと呼ぶ) という言葉に関連付けられる様々な IT 機器, システム, そしてサービスが実際に導入されつつある。携帯電話の普及と高機能化により, 電子メールはもとより, Web ブラウジング, 音楽のダウンロードや再生, そしてワンセグによる放送コンテンツの視聴などが可能となっている。但し, このような携帯端末を利用した IT システムはモバイル・コンピューティングと呼び, ユビキタスとは区別されるケースも多い。本来のユビキタスは, 人間の生活空間のあらゆる場所や環境に計算資源, すなわちコンピュータが埋め込まれている, という状況を指す場合が多い [1] [2]。

本稿では, シンプルに「いつでも・どこでも・だれにでも」簡単に利用できる情報インタフェースという観点で捉え, ユビキタスとモバイルとの区別の必要はなく, 「時間 (いつでも)・環境 (どこでも)・人 (だれにでも)」が多様性をもったときに, いかにその多様性に適応し,

利用しやすいインタフェースを提供するか, という点にスポットをあてる。すなわち, ここでテーマとするユビキタスなマルチメディア・インタフェースは, 次の3つの側面に重点を置く。

- 時間：いつでも利用しやすい。とくに時間がないとき・急いでいるときでも快適に利用できる, すなわち時間を有効に活用できるインタフェースを提供する。
- 環境：どのような環境・場所でも利用できる。ここでは特に我々の生活空間で大きなウェイトを占める公共空間における有益なインタフェースを提案する。
- 人：どのような属性をもった利用者でも快適に利用できる。ここではとくに, 情報リテラシの低い層の利用者に快適なインタフェースを提供する。

本稿では上記3つの側面に関するマルチメディア・インタフェース技術について紹介する。

まず, 時間という側面に関しては, 長いコンテンツを自動的に短くし, 短時間でのプレビューを可能とする自動要約技術「Choco Para」について記述する。近年は, 従来の DVD や TV などのメディア以外に, ネットワークを介して膨大な量のコンテンツにアクセス可能となっている。ブログや You Tube に代表される多くのサイトで自作の映像や様々な情報発信が可能となってきており, 視聴可能なコンテンツは爆発的に増加し続けている。これに対し, 人間の一日あたりの視聴時間は, 平均

* NTT アドバンステクノロジー株式会社
Meia Integration Business Unit, NTT Advanced Technology Corporation.

** NTT サイバーソリューション研究所
Cyber Solutions Laboratories, Nippon Telegraph and Telephone Corporation.

的には4時間程度であり[3], 将来的にも抜本的に増えることはないと考えられる。このようなトレンドにおいて, 短時間でコンテンツを「ざっとプレビューする」ことは, 非常に重要なヒューマンインタフェース (HI) となる。

環境に関する側面としては, ここでは, 公共空間における新しいメディアの創出を考える。我々の日々の生活において, 駅や空港, 電車や車による移動など, 公共的な環境にいる割合はかなり高い。しかしながら, 街頭における大型 LED 画面による映像表示のようなケースを除けば, 有益な情報提供メディアは現在のところ存在しない。また, 大型 LED 画面による映像表示は, TV などと同類で, 基本的に利用者側から操作は行えない一方向的なメディアである。本稿では, 大型の画面を用い, 多人数に同時に露出しながら, 人々の動きや環境の変化によってインタラクションが可能な新しいメディア「みらいチューブ」を紹介する。

人に関する多様性への適応という側面に関しては, 一つの社会問題にもなりつつあるデジタルデバイドに関係する。PC や携帯電話の普及によってインターネット利用者は増加し, かつ, 単なる情報のブラウジングからブログや SNS (Social Networking Service) のような情報発信, Wikipedia に代表される Web2.0 のような集合知の潮流など, その活用・利用方法はますます多様化・高度化しつつある。一方, もともと IT 技術になじみのない利用者層は, このトレンドのため, ますます心理的に高いハードルを感じ, 結果として両者の格差は助長される傾向にある。このような状況を鑑み, Web ページの閲覧など, 本来能動的に PC 操作を行わないと実行できないプロセスを, TV やラジオのように受動的でかつ娯楽性の高いメディアに変換する技術「 ∂ ロボ (デルロボ)」

を紹介する。概念的には Web ページの閲覧を, テレビのように受動的で簡単に操作が可能で, かつ娯楽性の高いメディアに変換するシステムを考える。この技術によって情報リテラシの低い層に対して, 簡便に Web ページにアクセスする手段を提供する。あるいは, TV の使われ方の多くがそうであるように, 何か他のことをしている状況でも, なんとなく Web 情報が流れていて, 興味のあるトピックがでてきたときに, すかさず注視する, という新しい Web 閲覧スタイルを提供する。

以下の章では, 「時, 環境, 人」に関連する上記 3 つの技術, すなわち, 公共空間での新しいメディア「みらいチューブ」, コンテンツ要約技術「*Choco Para*」, そして, Web 閲覧を受動的で娯楽性の高いものとするコンテンツ変換技術「 ∂ ロボ」について紹介する (図 1 参照)。最後の章では結論と将来の展望について触れる。

2. 公共空間の新しいメディア “みらいチューブ”

日々の生活の中で, 駅や空港, 電車やバス, あるいは公園や街頭など公共的な空間を訪れている時間は思いのほか長い。生活の場におけるユビキタスな情報システムとして, 篠原他[4]は, 駅・空港などの公共空間において, 有益でかつ楽しめる, 新しいインタラクティブなマルチメディア・インタフェース「みらいチューブ」を提案している。本章ではその概要と実証実験について紹介する。

2.1 みらいチューブのコンセプト

電車や駅のコンコースなどの公共空間における従来のマルチメディアシステムは, 通常一方的に映像を表示するものと, 利用者の何らかの操作やアクションによって操作可能なインタラクティブな端末が考えられる。前者としては街頭交差点などの LED 大型ディスプレイがあ

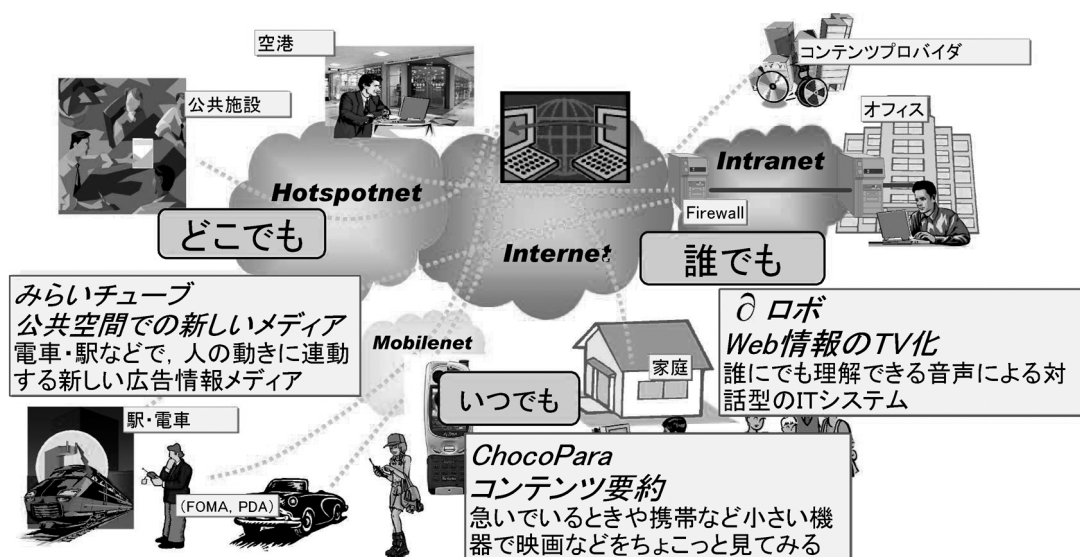


図1 「いつでも・どこでも・誰でも」を実現する3つの技術コンセプト

り、後者としては座席予約・各種案内など特定目的のためのパーソナルユース（同時利用者は一人）を前提としたキオスク端末の類がある。みらいチューブのコンセプトでは両者の中間的な情報メディア、すなわち巨大な画面で大勢が同時に閲覧することを想定し、かつ、利用者の何らかのアクションあるいは環境の変化によって表示内容が適応的に変化するようなインタラクションを実現する新しい情報メディアインタフェースを提案している。

みらいチューブの概念的な構成を図2に示す。通常システムとの違いは、複数のセンサからの入力を有機的に取り扱える点と、出力も複数の画面（及びスピーカ）を含んでいる点である。Situation Recognizerは、複数のセンサからの入力を解析・統合する。ここでセンサの入力はカメラからの映像情報やマイクからの音情報、あるいはRfID (Radio Frequency Identification) など無線系デバイスからのID情報など様々なメディア情報が想定されている。解析・統合された結果は、歩行者の位置や動き、音声認識結果、ID情報など意味的な情報に変換されInteraction Creatorに送出される。Interaction CreatorはInteraction Generation Ruleを参照し、入力情報とマッチするRuleがあればこれを実行する。たとえば「人が(x, y)に位置するならば⇒Icon-Aを(x, y)に対応する画面上の(x', y')にポップアップし、効果音Sを出力」のようなルールから構成される。

Media-Space Composerは、仮想的な3D空間を内部で保持し、Interaction Creatorによって生成されたインタラクション、すなわち何らかの表示情報を空間内に創り出し、複数のPlayerを通じて巨大な画面に表示する。仮想空間内には複数の2次元画像や3次元的な任意のオブジェクトを生成することが可能で、Playerを通じて複数のプロジェクトに画像が割り振られ、表示さ

れる。このように表示すべき世界と、それを物理的に表示するプロジェクトとは完全に独立した構成にすることで、実画面の大きさやプロジェクトの数に対する変更が柔軟に行えるスケーラブルな構造となっている。

2.2 みらいチューブ実証実験

みらいチューブのコンセプトを実フィールドで評価するため、実験システムを構築した。実験は、横浜高速鉄道みなとみらい線みなとみらい駅のB3Fコンコースで行った。改札を出て直後にある広大な半円筒型（Vault型）公共空間であり、1日に1000人程度の通行人がある。

実システムでは、図3にあるような12 m × 3 m（500インチ相当）の映像画面を半円筒状の壁に8台のプロジェクトによって投影する。カメラを含む実際の構成を図4に示す。

2.3 インタラクティブなパブリックメディア

みらいチューブの実証実験は、みなとみらい線が開業した2004年2月から10月末までの9ヶ月間実施された[4] [5]。この間、様々なデジタルアートや広告コンテンツ、映像による時報などを表示し、利用者に対して公共的な受容性（印象）、認知度、理解度、興味度などをアンケート調査によって評価した。

みらいチューブの大きな特徴は、通行人などの動きや位置を認識することで表示を変化させることができる点であり、この一種のインタラクティブ性によって従来の一方的に流すだけの表示システムに比べ、より興味や利便性を高めることを狙っている。図5に典型的な4種のインタラクションを示す。Type A, Bはいずれも利用者の動きをアイコンなどが追跡するタイプである。Type Cは丁度鍵盤の上を歩いているようなイメージで対応する鍵盤のキーの色が変わり、かつ「ド、レ、ミ」など音高の異なる効果音になる。Type Dは、画面が大

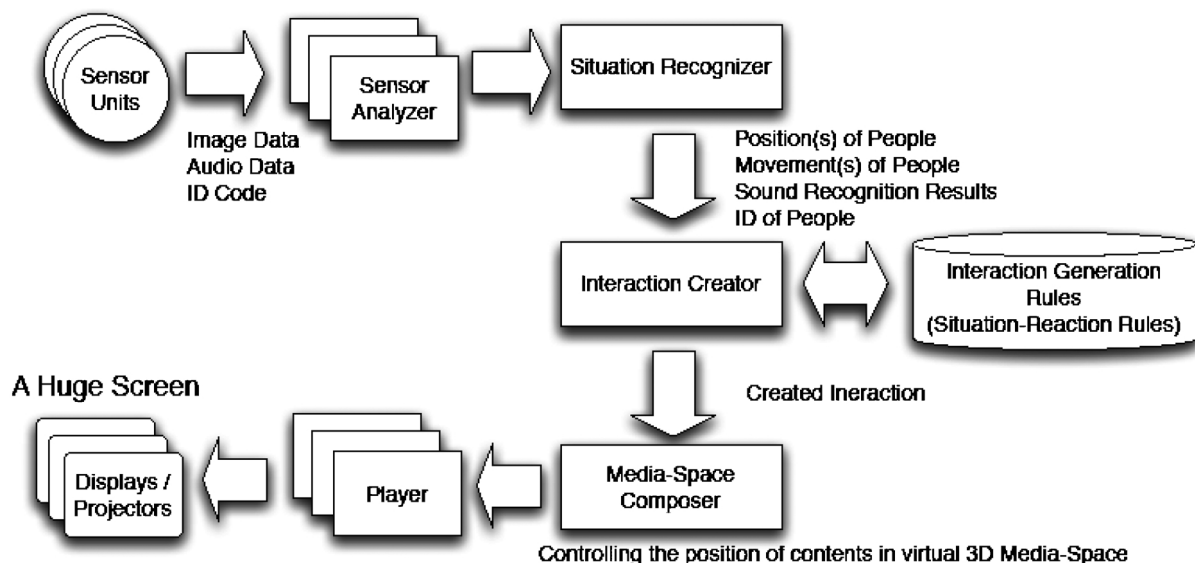


図2 みらいチューブの概念的構造

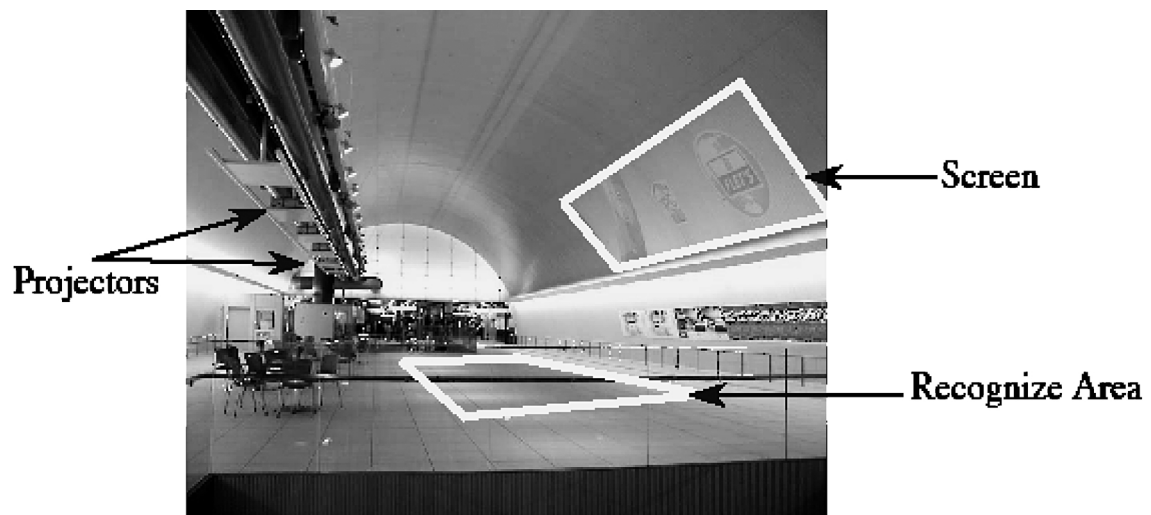


図3 みらいチューブ実験システム

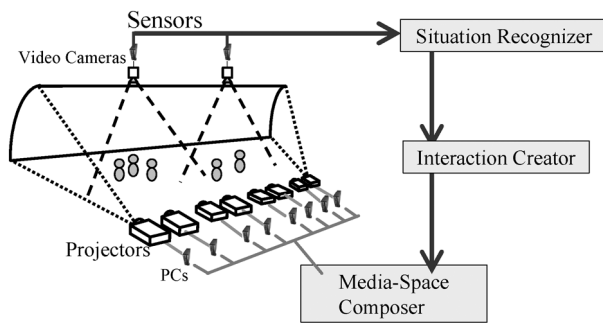


図4 みらいチューブ実験システムの構成

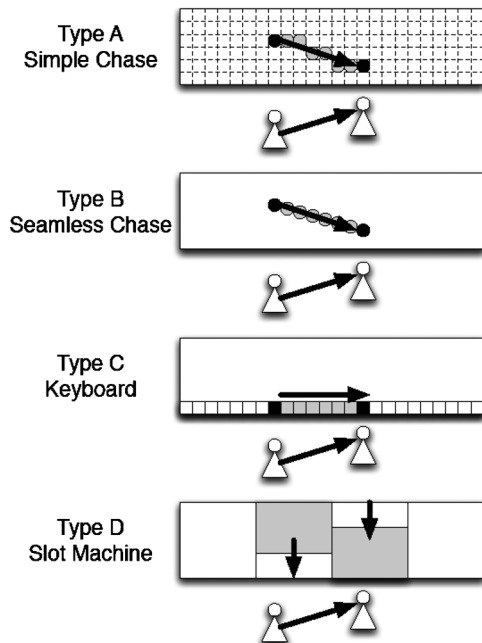


図5 みらいチューブにおける主なインタラクションのタイプ

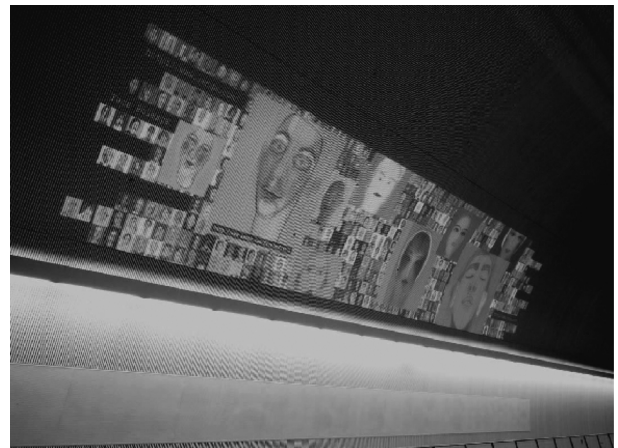


図6 デジタルアート“顔 Poiesis”（安斎利洋氏・中村理恵子氏による作品）

きく4つの領域にわかれており、対応する領域に入ると、スロットマシンのように当該区域の画像が回転するしくみとなっている。

表示するコンテンツとしては、種々の広告のほか、デジタルアートの展示も試みた。図6に示したのは「顔ポイエーシス」とよばれるもので、安斎利洋氏・中村理恵子氏の作品である。この作品では多数の人工的に創り出された顔が表示され、誰かが対応する位置に立つと、その顔が大きくクローズアップされるような工夫がなされている。

実証実験では、公共空間における受容性や広告メディアとしての目立ちやすさ・理解のしやすさをアンケート調査によって評価した。受容性はかなり好意的な印象をもたれたことが確認できており、またインタラクティブな動きによって多くの通行人に興味をもって迎えられたことなどが確かめられた。発展形態としては、人の動きや動作だけでなく、例えば天気のような外部環境の変化

を反映させる（雨・夕日・星空など）、あるいは、画像だけでなく、音の変化などをインタラクションに盛り込むなどの拡張が考えられる。

3. 映像コンテンツを自動要約する技術 “Choco-Para”

ブログや SNS (Social Networking Service) が普及し、YouTube に代表される所謂 CGM (Consumer Generated Media) が一般化しつつあり、我々が閲覧可能なマルチメディアコンテンツは爆発的に増えつつある。一方、一人の人間の視聴に費やせる時間は一定であり、モバイル環境における視聴機会は増えつつあるものの、著しく増加することはない。また、ワンセグなどモバイル環境での映画・TV などの視聴は実現されつつあるが、利用形態から考えて、長時間の視聴は考えにくく、限られた時間での視聴が殆どであろうと推定される。

このようなトレンドを鑑み、日高他 [6] は、映画・ドラマ・スポーツなど様々なコンテンツを音声の強調度を基に自動要約し効率的に視聴する新しいインタフェース Choco Para を実現した。本章ではその概要と現時点での性能について紹介する。

3.1 コンテンツ要約技術の概要

まず、本稿においてマルチメディアコンテンツを要約する、あるいは効率よく視聴するインタフェースとはどのようなものか定義しておく。ここでは、そのコンテンツにおいて最も興味を喚起する部分のみを抜き出し、これを提示することによって、「要約」または「視聴効率の良い」視聴インタフェースであると考え。また、「最も興味を喚起する」部分の抽出は、現時点では「音声的に最も強調されている」部分の推定によって実現する。本来は映像情報や意味的な内容を汲み取り総合的に推定すべきであるが、対象とするコンテンツのジャンルや種類を特定せず汎用的に利用できる技術を想定すると、映像情報や状況の正確な把握は極めて困難であり、現実的な解として音声の強調推定によって要約を構成するアプローチをとる。

このような考え方を基本においたコンテンツ要約処理の概略を図 7 に示す。

図 8 には、上記処理の実行の様子を概念的に示した。図 7 に示したように、マルチメディアコンテンツからまず音声情報が抽出され、韻律パラメータ（声の高さ・強さ・速さ）をもとに、予め学習された強調確率テーブルから当該箇所がどの程度強調されているかという強調度を推定する。強調度は強調されている確率に相当する尺度である。一般に強調されている語句に続く単語や文節も同時に強調の対象とされており、言語的な連続性を考慮して、強調部分を含む音声における「段落」に相当する部分を抽出する（これを音声段落と呼ぶ）。図 8 では、P/Q/R/S が強調度の高い部分であり、これ

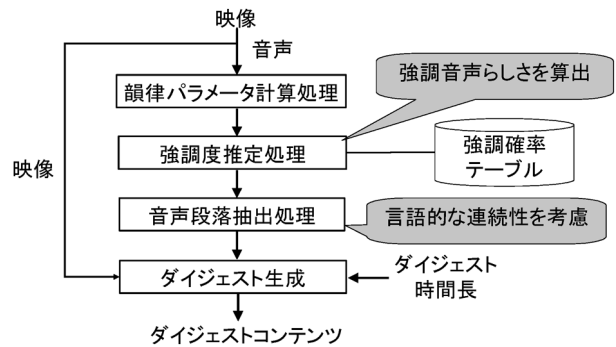


図 7 コンテンツ要約処理の概要

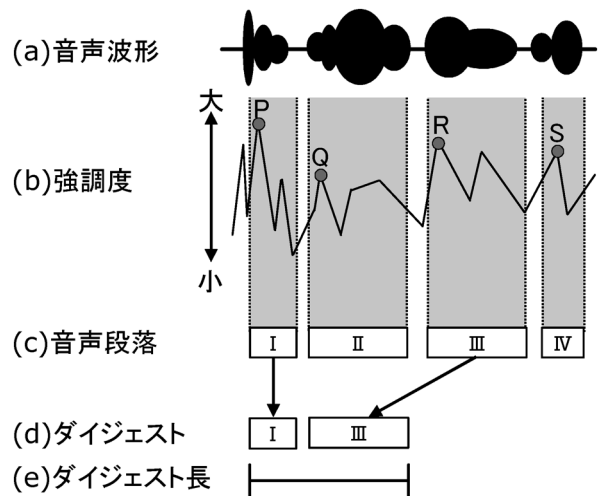


図 8 コンテンツ要約の概念的処理例

を含む音声段落が I/II/III/IV となる。換言すれば音声段落 I/II/III/IV の強調度はその内部で最も強調度の高い数値、すなわち PQRS が代表値となる。予め利用者などから設定されている要約率、すなわち最終的な要約コンテンツの尺をもとに、強調度の大きい音声段落に相当する映像を順次接続することによってダイジェストコンテンツを生成する。図の例では、I と III が他の音声段落よりも強調度が高く、かつ指定された要約コンテンツの尺を超えないように、これら 2 つの音声段落が選択されている。

以上のような要約方式の主な特長は次のようにまとめられる。

- ① 音声認識技術を利用しておらず、雑音耐性に優れた韻律パラメータ分析を基にした処理であるため、一般のホームビデオから BGM などの入ったドラマ・映画など幅広く利用可能となっている
- ② 強調度の大きい音声段落から選択していくという動的な方法によって、任意の要約率を利用者が設定できるという極めて柔軟なインタフェースを実現している。
- ③ 韻律パラメータは言語依存性が低く、多言語に対

する適用も可能である。

3.2 強調度推定方式

本説では韻律パラメータを基にした、強調度の推定方式について述べる。強調度を推定するにあたり、次の2点は重要なポイントである；1) 強調は韻律パラメータの時間的な構造・遷移によって推定されるべきである。2) 強調された音声特徴はある種のゆらぎや幅を含んだものとなっている。本方式では、前者については量子化された韻律パラメータの遷移情報を取り入れることで、後者については、強調性から平静性（非強調性）との差分によって強調度を定義することで、各々対処している。

まず学習用の音声データに人手によって「強調部分」と「平静（非強調的）部分」をマーキング（ラベリング）を行う。韻律パラメータをベクトル量子化 [7] し、あるベクトル C_f に関して、当該ベクトルが強調である確率 $P_{emp}(C_f)$ 、 C_{f-1} から C_f に遷移したときに強調である確率 $P_{emp}(C_f|C_{f-1})$ 、および、 C_{f-2} から C_{f-1} そして C_f に遷移したときに強調で有る確率 $P_{emp}(C_f|C_{f-1}C_{f-2})$ を全てのベクトルの組合せについて計算しておく。同様に、平静に関しても $P_{nrm}(C_f)$ 、 $P_{nrm}(C_f|C_{f-1})$ 、 $P_{nrm}(C_f|C_{f-1}C_{f-2})$ を学習データより求めておく。

音声強調推定時には入力音声の韻律パラメータをベクトル量子化し、当該音声フレーム f の強調確率 $P_E(f)$ 平静確率 $P_N(f)$ を (1) (2) 式によって求める。

$$P_E(f) = \lambda_{e1}P_{emp}(C_f|C_{f-1}C_{f-2}) + \lambda_{e2}P_{emp}(C_f|C_{f-1}) + \lambda_{e3}P_{emp}(C_f) \quad (1)$$

$$P_N(f) = \lambda_{n1}P_{nrm}(C_f|C_{f-1}C_{f-2}) + \lambda_{n2}P_{nrm}(C_f|C_{f-1}) + \lambda_{n3}P_{nrm}(C_f) \quad (2)$$

ここで λ_{ei} , λ_{ni} ($i=1, 2, 3$) は重みであり削除補間法 [8] によって求める。 L フレームからなる音声部分 X の強調確率・平静確率は (3) (4) 式で求める。

$$P_{Xemp} = \prod_L P_E(f) \quad (3)$$

$$P_{Xnrm} = \prod_L P_N(f) \quad (4)$$

強調確率・平静確率の値から、音声部分 X の強調度 K_X は、強調確率と平静確率の対数の差を時間長 L で正規化した次式によってもとめる。

$$K_X = \frac{\log P_{Xemp} - \log P_{Xnrm}}{L} \quad (5)$$

このように、強調度を強調らしさから平静らしさを差し引くアプローチによって、たまたま強調度が実際より過大評価されてしまう危険性を回避し、より安定した強調度推定を可能にしている。また強調らしさ・平静らしさは量子化されたベクトルのトライグラム（三項連鎖）によって表現されており、韻律パラメータの時間的な遷移構造を反映したものとなっている。

3.3 有効性の評価と将来への展望

Choco Para 方式による要約コンテンツの有効性を確

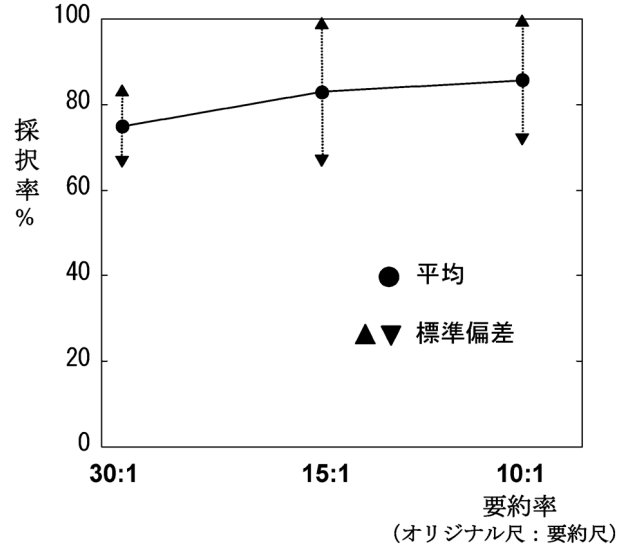


図9 Choco Para 方式と等間隔選択方式の対比較試験結果

認するため、10人の被験者に対し、7種の映像コンテンツについて本方式と、一定間隔による間引き方式のどちらのほうが要約として好ましいかという対比較実験をおこなった [9]。なお、要約率は10:1, 15:1, 30:1の3種について評価した。結果を図8に示す。

図からも分るように、いずれの要約率においても80%程度の高い採択率であり Choco Para 方式が等間隔方式より優れていることが分る。

コンテンツ要約技術は、広義には感性的な情報による情報検索・閲覧インタフェースの範疇であり、今後ますますニーズは高くなっていくと考えられる。ひとつの発展系としては、感性情報、すなわち「悲しい場面」「楽しい場面」「元気な場面」などの検索やプレビューなどへの適用が試みられている。[10]

4. Web 情報を TV 化するインタフェース “@ ロボ”

近年インターネット利用は増加の一途をたどり、携帯電話によるアクセスも含めれば、いまや子供から高齢者まで幅広く広まりつつある。しかし、高齢者をはじめ、まだインターネットへのアクセスに抵抗があり、PCを利用する機会の殆どない情報リテラシの低い層のあることもまた事実である。利用する機会を持った人々は、さらに高度なユーザとなる可能性があり、所謂デジタルデバイドは深まる一方ともいえる。全世界の Web 情報の閲覧インタフェースに TV と同程度の「便利さ、手軽さ、面白さ」を提供することによって、情報リテラシの低い人々に対するアクセシビリティ向上を図る。

このようなコンセプトから菊地他 [11] は、本来能動的である Web 情報へのアクセスを TV と同等の簡便なインタフェースに変換をすることにより、受動的な視

聴を可能とするインタフェースを提案している。ここで、テキスト・写真などが主となる「静的」なWeb情報を、TVのような「動的」な情報に変換・再構成する際には、様々な演出のヴァリエーションやアプローチが可能である。携帯電話やホットスポットの普及など、情報システムの利用シーンが多様多様であることを受けて、我々は、利用環境や利用者の心的状態に最もマッチした演出・プレゼンテーションを自動的に生成する手法を提案する。ここで、TVニュースなどの場合はキャスターやアンカーと呼ばれる人たちによって視聴者に興味深く情報提供していることのアナロジーから、Web情報内容を身振りと音声によって紹介するロボットの導入を試みている。このようなガジェット(道具立て・小物)は、利用者の興味を引き、面白さの向上などに非常に有益であると考えられる。なおロボットは必要に応じて入退場するように実現されており θ (デル) ロボと呼ばれている。

4.1 感性情報に適合したプレゼンテーション生成

IT利用環境のユビキティ(遍在性)は向上してきており、利用可能なサービスも多様化してきている。またこれに呼応する形で、利用目的や利用者の置かれている状況・心理的な状態も多種・多様なケースが考えられる。

そこで、我々はこのような利用状況・状態の多様性を念頭におき、これに最もマッチした形でWeb情報を再構成し利用者に提示するモデルを提案する(図10参照)。まず、コンテンツ再構成において参照すべき情報システムの利用状況を、利用者状況空間(User's context space)とよぶ。利用者状況空間は、アクセスしているコンテンツの属性、利用者の情報リテラシと心理的状态、利用している端末や場所・時間、の3つの項目によって特徴付けられる。利用者状況空間は、このままの形ではコンテンツ再構成に利用しにくいいため、よりプ

レゼンテーション生成・演出に近い情報、視聴要件空間(Looking and listening requirements space)にブレイクダウンされる。視聴要件空間は、感性情報、コンテンツサイズ(尺、画面などの大きさ)、理解度、の3つの要因で構成される。我々の最終目的とするインタフェースシステムは、ターゲットであるWebコンテンツを視聴要件空間に最もマッチした形で再構成し、プレゼンテーションを生成する。

上記のインタフェースシステムを実現するためには、利用者状況空間からの視聴要件空間へのブレイクダウン、視聴要件空間に基づく、コンテンツ再構成、などを実現する必要があるが、本技術ではその第一ステップとして、視聴要件空間における感性情報からのプレゼンテーション生成について実現した。

このモデルに基づき、Webコンテンツと利用者の心理状況に基づくプレゼンテーション生成の処理フローを図11に示す。ここで利用者の心理状況とは、より正確には、「利用者に喚起すべき感性(感情)」を指し、例えば「楽しい・嬉しい」「軽やかに」「悲しげに」などの感性語で表現されるものを想定する。

ターゲットとなるWebコンテンツは、コンテンツ要素抽出部(Content Element Extraction)に渡され、コンテンツのカテゴリ、タイトル、内容テキスト、写真、図などの属性要素に分類され、プレゼンテーション再構成部(Presentation Composition)に渡される。プレゼンテーション再構成部では、テキストの音声化、テロップの生成、視覚的な効果の付加、BGMの付加、 θ ロボの動作設定、などを行う。これらの演出や効果は、目的とする感性情報をもとに、状況・再構成文法(Situational Composition Grammar)を参照することによって、決定される。再構成された表現はプレゼンテーションインタフェース部に引き渡され、利用者に提示され

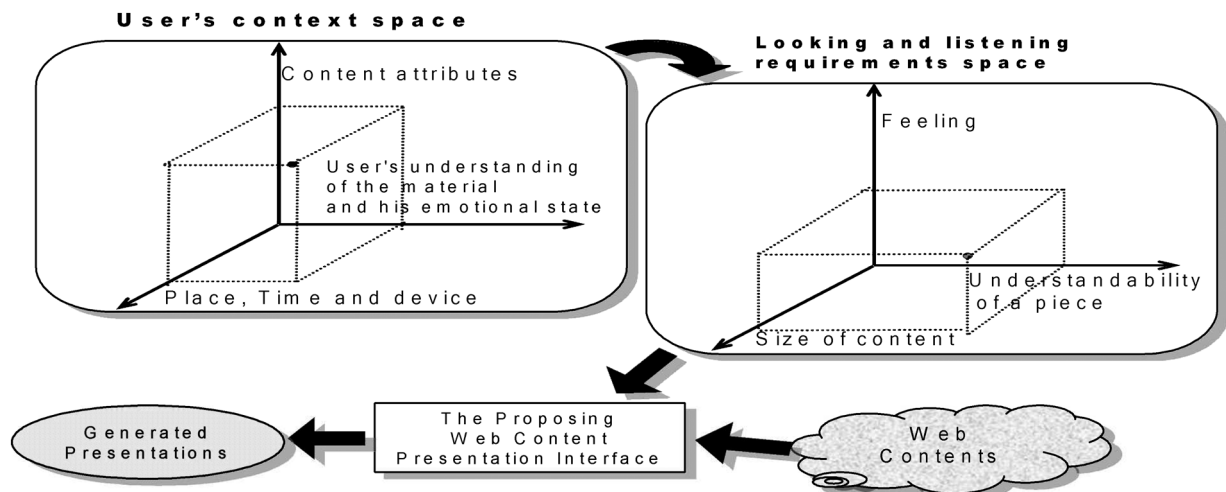


図10 利用者の状況を考慮したコンテンツ再構成モデル

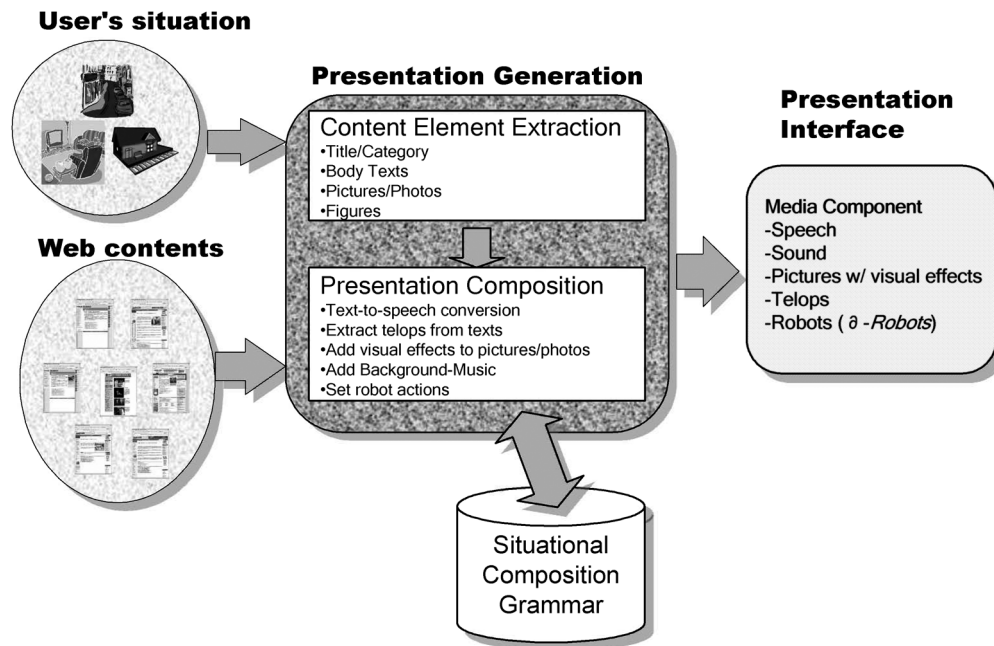


図11 感性情報に基づくプレゼンテーション再構成インタフェース

る。プレゼンテーションインタフェースでは、音声・音、テキスト・テロップ、図・写真、 θ ロボ、などのメディア要素（Media Component）に応じた表出デバイスが使用され、利用者に再構成されたプレゼンテーションを表示する。

4.2 感性を反映したコンテンツ再構成のアプローチ

一般に Web コンテンツは「タイトル」「本文」「画像」「画像を説明するテキスト」[12] の4つの部分から成り立つことが多い。具体的な Web 情報から TV 型コンテンツを再構成する手順は下記のようなものとなる。

- タイトルや画像を説明するテキストのキーワードを使ってテロップを作る [13]
- 静止画を使って動画を模したものを作る
- 本文を使って読み上げ音声を作る
- 読み上げ音声に応じて適切なロボットのアクションを構成する
- 必要に応じて適切な BGM をつける

ここで、テロップや静止画の表示の仕方や BGM、ロボットの有無や振り付けなどによって、視聴者のうける印象はかなり変わってくると考えられる。

感性（印象）をもとに最適な演出効果を得るために、イメージや印象情報を取り扱う場合に有効な SD 法（Semantic Difference）[14] を用いる。SD 法では感性語対（暖かい⇔冷たい、堅い⇔柔らかい、など）を利用して、対象を主観評価させ因子分析を行うことにより、印象空間における感性語対と評価対象の布置を得ることが出来る。図12に主な感性語の印象空間における布置を、図13に主な効果の布置を示す。印象空間の第1の因子（寄与率48%）は「活発（Active）⇔静寂（Calm）」

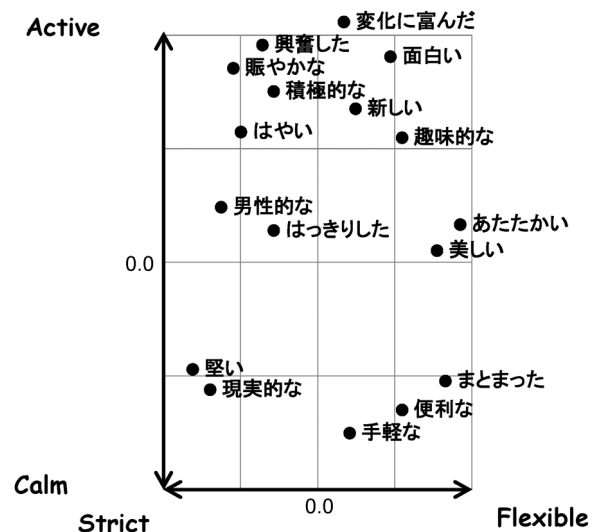


図12 SD 法因子分析による印象空間における感性語の布置

の軸であり、第2の因子（寄与率%）は「柔軟（Flexible）⇔堅固（Strict）」と表現される特徴となった。また、効果の中では、画像効果の「スポットライト、ズーム切り替え・速い」などが「活発」であり、「ズーム+パン」などは「柔軟」となっている。また θ ロボは、これらプリミティブな効果と同じ座標では評価できないが、活発かつ柔軟である方向に印象をシフトすることが主観評価実験によって確かめられている。

これらの結果をベースとして、ターゲットとなる印象をもとに、Web コンテンツを再構成する方法を図14に示す。利用者（視聴者）に与えたい印象 x を入力として、最も類似性の高い感性語の印象空間座標（ I_x, I_y ）

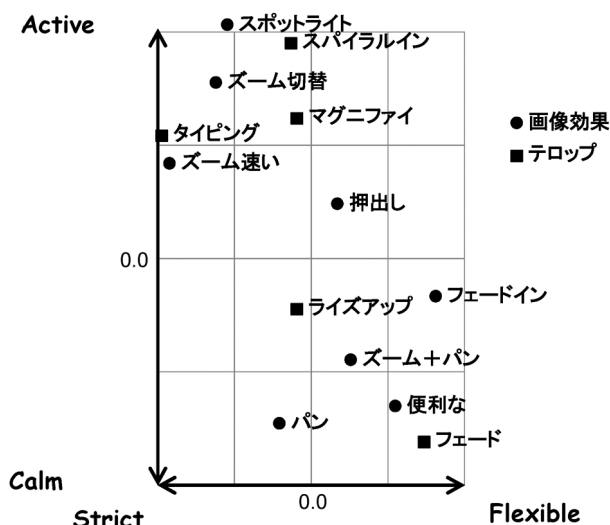


図13 SD 法因子分析による印象空間における各種効果の布置

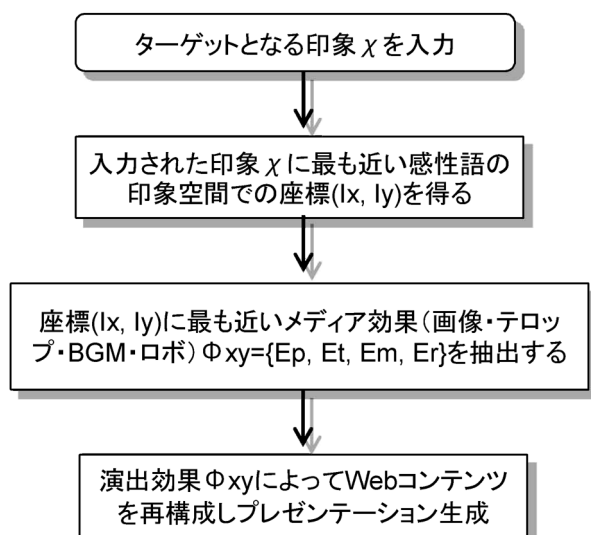


図14 感性情報に基づく Web コンテンツ再構成のアプローチ

を得、これをもとに最適な演出効果の集合 $\Phi_{xy} = \{Ep, Et, Em, Er\}$ を設定する。最終的にはこれらの効果をもとに Web コンテンツの再構成を行う。

4.3 Web コンテンツ TV 化の効用と展望

実験システムを試作し (図15), 上記のようなアプローチによっていくつかの Web コンテンツを再構成し, 主観的な評価を行ったところ, 直感的に興味を引き, かつ内容の理解しやすさの助けになることが明らかとなっている。利用者に与える印象がどの程度適切に設定できるのか, 個人性やコンテンツの内容とのマッチングはどのように関係してくるのか, などについては今後検討を進める必要がある。デジタルデバイドの問題が顕在化する傾向の中で, このような利用者の感性に直接訴えていく, 利用し易いインタフェースは, 将来の IT システムにおいて, ますます重要な役割を担っていくと考

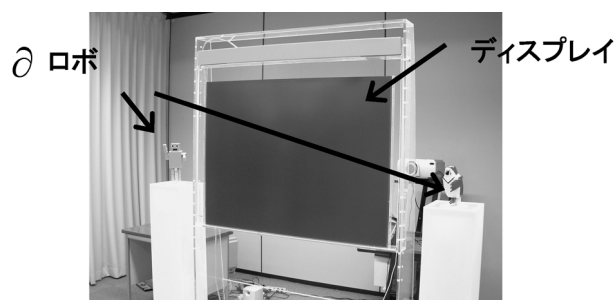


図15 d ロボ実験システムの外観

えられる。

5. 結 言

ユビキタスなマルチメディア・インタフェースとして, 「時間 (いつでも), 環境 (どこでも), 人 (だれにでも)」利用しやすいインタフェースという視点で捉え, 膨大なコンテンツを短時間で効率よく視聴可能なコンテンツ要約技術 “Choco Para”, 公共空間における新しいインタラクティブな巨大画面メディア “みらいチューブ”, および, 情報リテラシの低い層に対する Web コンテンツのアクセシビリティを向上する高度な TV 化変換インタフェース技術 “d ロボ” について紹介した。いずれの技術・コンセプトは大きな拡張性を秘めており, より高度で汎用性・利便性の高い技術へとブラッシュアップが可能である。

コンテンツ要約技術では, すでに感性情報によるコンテンツ検索へと拡張が試みられている。この方法論をさらに発展させれば, ソースである映像コンテンツを, 各個人の感性によって自動編集・チューニングされた “My Own プレビュー” として再構成することも可能である。また, この編集・チューニングの対象は複数の多くのコンテンツを同時に対象とすることができる。即ち, “過去一年間で製作された映像コンテンツから感動的な場面で3分のプレビュー作成” というようなリクエストも可能である。無論, 著作権・改変権に関しては注意を払う必要があるが, “映像コンテンツの感性によるコラージュ” が再生成でき, 新しいコンテンツ視聴スタイルの萌芽する可能性も有ろう。

公共空間におけるインタラクティブ・メディアに関しては, 前述のように人々の動きの他に, 外部環境・自然現象の変化に適応するインタラクションへの拡張も可能であろう。地下空間なので, 例えば, 雨や風がひどくなってきたときに, それを示唆する映像や音を表現するということもできる。単なる広告メディアではなく, 公共空間における時間や環境・場所, そして人の流れなどに適応的に動作する新しいインタラクティブ・インタフェースとして高い拡張性を含んでいる。

Web 情報の TV 化変換に関しては, 現時点では, あ

る目標とする感性情報が与えられ、これに適した演出はどのように構成するか、という点に重点をおいてコンテンツの再生成を行う方法について検討した。将来的にはコンテンツの内容や、その人の置かれている状況なども考慮にいった総合的コンテンツ再生成へと拡張していくことが可能である。すなわち、環境・コンテンツ内容・利用者個性・状況などから、もっとも適したコンテンツ再構成を行うインタフェース技術の確立という目標が想定される。恐らく、このような情報変換端末は、各個人に専用である必要があり、現在の携帯電話のように常に持ち歩いていて、自動的にNWに“Plug-In”し、置かれている状況に最適な方法で情報やコンテンツを表出するようなスタイルで利用されると思われる。

参 考 文 献

- [1] 坂村健, “ユビキタス・コンピュータ革命”, 角川書店, 2002.
- [2] Mark Weiser, “Some Computer Science Problems in Ubiquitous Computing,” *Communications of the ACM*, 1993.
- [3] “デジタルコンテンツ白書2002”: 経済産業省商務情報政策局監修, 財団法人デジタルコンテンツ協会編, 2002.
- [4] Shinohara, Tomita, Kihara, Nakajima, Ogawa, “A huge screen interactive public media system: Mirai-Tube”, *Proceedings of HCI2007*, pp936-945, 2007.
- [5] 篠原, 富田, 木原, 中畠, 小川, “公共空間における巨大インタラクティブシステム—みらいチューブ—”, ヒューマンインタフェース学会論文誌, Vol. 9, No. 3, pp. 305-312, 2007.
- [6] 日高, 竹内, 松浦, 茨木, 中畠, “音声の強調情報を利用したビデオコンテンツ短縮試聴方法の検討”, 画像電子学会 Vol. 34, No. 5, pp505-511, 2007.
- [7] Linde, Buzo, Gray, “An algorithm for vector quantizer design”, *IEEE Transactions on Communication*, 28, pp84-95, 1980.
- [8] Jelinek, Mercer, “Interpolated estimation of Markov source parameters from sparse data”, *Pattern recognition in practice*, Amsterdam Elsevier, pp. 381-397, 1980.
- [9] Hidaka, Nakajima, “A new multimedia content skimming technique at arbitrary user-set rate based on automatic speech emphasis extraction”, *International Journal of Human-Computer Interaction*, Vol. 23, No. 1 & 2, pp. 115-129, 2007.
- [10] 入江, 日高, 佐藤, 谷口, 中畠, “CGM 動画を対象とした感情表出区間自動検出法”, 電子情報通信学会全国大会論文集 D-12-94, 2007.
- [11] Kikuchi, Hidaka, Nakajima, Kobayashi, “Chat-Robot Based Web Content Presentation Interface and Its Evaluation”, *Proceedings of HCI2007*, pp. 934-943, 2007.
- [12] Chakrabarti, S., et al.: Automatic resource compilation by analyzing hyperlink structure and associated text, *Proc. of 7th World Wide Web Conference*, pp. 65-74, 1998.
- [13] 廣嶋, “統計手法に基づく Web ページからのヘッドライン生成” 自然言語処理 Vol. 149, No. 7, 2002.
- [14] 岩下豊彦 “SD 法によるイメージの測定”, 川島書店, 1983.