

図3 コンテンツの選択・検索が困難な時代へ

また、コンテンツ製作側から見た場合も、情報量は指数関数的に増大してきている。コンピューターで多くの処理を自動化するとはいうものの、やはり最終的に芸術として作品を仕上げていく過程はその人の創造性が重要になる。よく言われている事として、2009年には1人当たりの年間情報消費量は大体 10^{23} bitぐらいになる。[1]すなわち、人間の脳の記憶数を超えるぐらいの情報量が人間に入ってくる時代になる。このような時代に、人間がそれぞれの創造性を発揮するために、どうすべきかが非常に大きな課題になる。創造支援技術という視点で言えば、一人ひとりが簡便に芸術文化を創出できるような環境を整備することにより、国や民族を超えた芸術文化の交流を活性化するというのが1つの方向性であると考えられる。また、発信・利用技術の視点では、創造したコンテンツを双方向で意のままに利用するための流通・検索の仕組みをつくることが挙げられる。

これらの環境や仕組みの実現には、人間の視聴覚、触覚、言語、知能、柔軟な制御等の知的な機能を人工的に実現するような技術や、人と機械、機械/電子情報網(ネットワーク)を介しての人と人の意思疎通(コミュニケーション)の自動化、円滑化を図る技術、更には情報の流れと実際の物の流通や機器の制御等、実世界の動きをつなぐ技術の開発が必要となる。

3 - 3 メディア文化資源の創造を支える最新技術動向

メディア文化資源の創造・発信・利用とこれを支える情報通信技術をマトリックスで示す(図4)。以下では、図4の赤字で示した実例について、創作支援技術、発信・利用支援技術、基盤技術の順に説明する。

メディア文化資源	用途	創作支援	発信・利用支援	その他基盤技術
リアルな表現 デジタル 動画映像 マンガ デジタルグラフィクス アニメーション ゲーム クリエイティブな表現		画像合成、CG技術 レンジファインダー 画像編集技術 ボイスタグシステム	携帯型電子端末 配信制御 コミュニティ生成支援	著作権保護 大容量記録メディア 高速通信回線
		色・質感再現技術 色計測値空間 による色再現	MPEG7 適応型動画配信 検索技術 対話型インターフェース	多言語間 コミュニケーション 多言語音声翻訳
		VR、AR	感性表現技術 高臨場感再現技術	技能・知能の 形式値化 認知発達発現研究

図4 メディア文化資源の創造・発信・利用とこれを支える情報通信技術

(1) 創作支援技術

レンジファインダー[2] [3]

レンジファインダーは、簡単に言えば3Dカメラである。ストロボを利用した非常にシンプルな原理により、撮像した画像から、平面画像に奥行き情報を加えた3次元の画像データを素人でも簡単に作成できる。これにより、撮影した被写体の大きさの取得や、他の画像との重ね合わせの表現が可能になる。以下に3D撮像原理を説明する。

図5に示すように、レンジファインダーは、普通のデジタルカメラにストロボが二つ付属したものとなっている。二つのストロボから異なる陰影を持った光のパターンを順次被写体に当てて撮像する事により、一つのピクセルに対し2種類の濃淡情報を得る。この2つの濃淡情報を用い、画像反射率との関係から一個一個のピクセルのカメラからの距離を計算する事により、三次元データを得ることが出来る。例えば、ソファの上に人形を置いて撮影すると、通常の色画像と同時に奥行き静止画像(距離画像)を撮影できる。距離画像はカメラからの距離を表現した画像であり、これを利用する事によ

り、図5に示すように人形の部分のみを簡単に切り出すことができ、例えば別に撮影した背景との合成や、3次元のコンピューターグラフィックス（Computer Graphics）の立体画像の自動生成が可能になる。

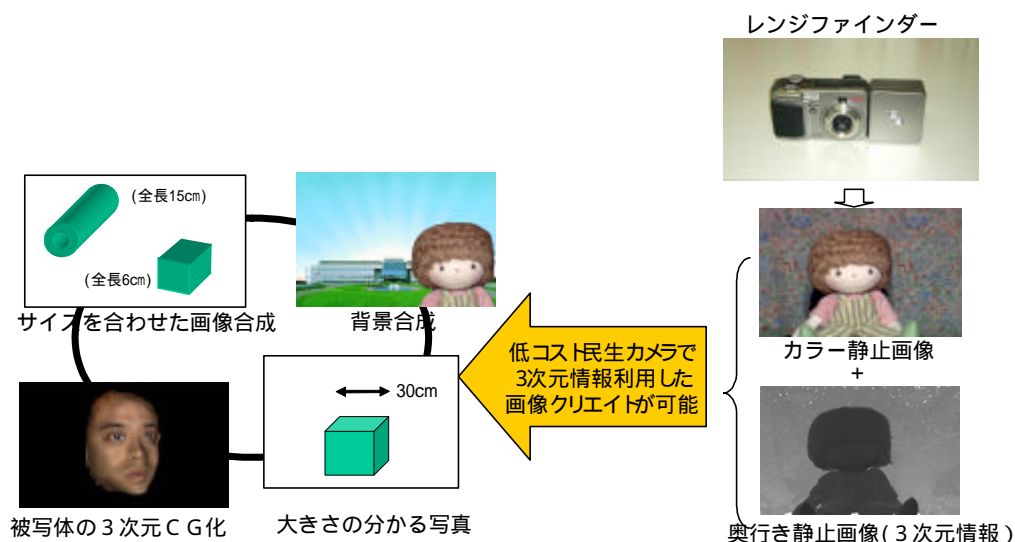


図5 レンジファインダーの原理と応用例

ボイスタグシステム[4]

映像や音声などのデジタルコンテンツを編集する為には、コンテンツに内容を示すタグを付ける作業が伴う。現在この作業は、編集者がカメラマンが撮影してきた映像を見ながら、主要な映像シーンに付与する適当なキーワードを決定し、アシスタントがキーボードで入力することによって行われている。例えば、キーボードから「ニューヨーク」というキーワードを入力すると、現在表示されている映像に「ニューヨーク」というタグがデジタルコンテンツの中に埋め込まれる。編集時には「ニューヨーク」と入力することで、「ニューヨーク」というタグが付与されたすべての映像を比較的簡単に取り出すことができる訳である。しかしながら、コンテンツ数が増加する中、この作業は正に非常に大変な作業になりつつある。この作業効率の画期的な向上と、作業に伴う疲労を飛躍的軽減を狙った新しい編集システムが、ボイスタグシステムである。

このシステムは音声認識技術を応用したもので、ディレクターが映像を見ながら、タグとして付けるべきキーワードを発声するだけで、このキーワードを自動で認識、テキストのタグにして映像に付与するものである。ボイスタグシステムを利用すると、編集時の疲労を大幅に低減出来ると共に、編集時間を従来の5分の1程度（従来20時間程度の編集時間を数時間程度）まで短縮、高速化できる。

色計測値空間における色再現技術[5] [6]

様々な文化財の保護という観点から、デジタル映像として納めた文化財の色をどのようにして現実の色に近づけるかという色再現の課題がある。様々な文化財をデジタル映像として保護していくという考え方は、電子資料館（デジタルアーカイブ）という観点でも非常に重要になってくる。通常、デジタルカメラなどで撮像した映像では、搭載されたイメージセンサーのデバイス特性に応じてRGB²の出力が出てくるため、そのままの状態でも保存しても、再現時に利用するCRT³の特性やメディアの特性が異なると、オリジナルに忠実に再現できないという問題が起こる。

この問題を解決する方法として、RGBからLAB⁴という色の計測値空間に写像して、文化財が本来持つ物理的な色の特性のパラメーターに変換して保存する色計測値空間を利用した色再現技術が研究されている。LAB形式で映像を保管しておく、再現時にテレビのCRTならCRTの特性に合わせて、LABから表示デバイスの特性に合わせたRGBに変換して再現することができる。このため、この色再現技術を利用して文化財を撮影しておけば、現物を見るのとほとんど変わらない正確な映像を半永久的に保存することができるようになる。

(2) 発信・利用技術

コミュニティ生成支援技術[7] [8] [9]

「本当に自分の見たい映画がどこにあるのか」、「自分が見たい絵画がどこにあるのか」を、例えばインターネットで探そうとするのは非常に困難な作業である。一般に趣味性の高いコンテンツにアクセスするには、趣味を同じくした人が集まる共同体（コミュニティ）での情報取得が効果的であると言われており、例えば、一般の地上波放送番組ではあまり放送されない特定のジャンルのコンテンツを大量にあつかうCSデジタル放送も一種の共同体（コミュニティ）といえる。

² RGB：電子機器で色を表現する際に用いられる表記法の一つ。色を赤(R)・緑(G)・青(B)の3つの色の組み合わせとして表現する。

³ CRT：電気信号を光に変換し、人間の目に見える像を発生させる表示装置。「ブラウン管」とも呼ばれる。

⁴ LAB：電子機器で色を表現する際に用いられる表記法の一つ。色を明度(L)、赤から緑への色度(a)、青から黄色への色度(b)により表現する。Lab色空間における変化量は、その変化によって受ける視覚の色変化の印象と比例する。

しかしながら、通信 / IT 技術の発展により、インターネット上などに構成される共同体（コミュニティ）の数は膨大になり、自分に合った共同体（コミュニティ）を見つけ出すこと自体が困難になりつつある。このような課題に対して、デジタルコンテンツの利用履歴を利用してユーザーの好みをあらかじめ推測し、個々のユーザーに合わせて仮想的な共同体（コミュニティ）を動的に作成する技術が研究されている。従来、ユーザーが自分に合った共同体（コミュニティ）を選択するという形であったものを、ユーザーが意識することなく、ユーザーを中心に、動的、自動的に非明示的共同体（コミュニティ）を生成する事ができる。これにより、常にユーザーの好みに合った映画や絵画展の情報のやりとりが可能となり、必要な情報に簡単にアクセスできるようになる。

MPEG - 7を用いた適応型動画配信技術[11] [12]

MPEG - 7という記述フォーマットを利用すると、ユーザーの好みに合わせた映像を即座に送信する事ができるようにもなる。MPEG - 7は既に標準化が進んでいる技術であり、MPEG - 7のマルチメディアコンテンツ記述に、コンテンツの内容説明や代表画像の情報を含めておく事により、ユーザーは映像コンテンツの最初から最後までを通してみなくても、コンテンツ内容や代表画像を簡単に検索可能となる。映像コンテンツそのものの選択だけでなく、更には映像コンテンツの中から見たい部分を見たい順番で自由に選択して見ることもできる。共同体（コミュニティ）生成技術などと合わせて適応型動画配信技術を利用すると、膨大な映画コンテンツの中から、ユーザーが見たいシーンをオンデマンドで取り出して見るということも実現可能になろう。

対話型検索インターフェース技術[13] [15] [16]

ユーザーには自分が見たい好みの映像や画像があっても、多量の画像や映像コンテンツの中から「これが見たい」というのを探すのは大変な作業である。これまでは、リモコン等でメニューを追って順繰りに探すというのが通常だったが、もっと簡単に、好みのコンテンツを探したいというニーズが拡大してきている。人と人がコミュニケーションする場合には、1対1の会話が成り立つことにより、スムーズに意思伝達できる。近年では、機械の側を賢くすることにより、同じことを人間と機械との間でできるようにする試みがある。

「機械と会話するとは如何なものか」という意見もあるが、むしろ前向きにとらえるならば、機械の側がきちんと人の意図を酌み取ってくれるような仕組みがあれば、デジ

タルデバイス⁵の問題を含めて、いくつかの重要な問題を解決することができる。ここでは、対話をしながらコンテンツを検索する技術について例を挙げて説明する。

図6は、音声認識リモコンによるデジタルテレビ番組視聴の概要を示した図である。現在のデジタルテレビは、従来の地上波放送に加えて、BSデジタル放送、CSデジタル放送を受信可能であり、非常に多数の番組を視聴可能である。一般的なのTVのリモコンを利用すると、例えば、サッカーの番組を見たい場合、メニュー画面からジャンル検索に入り、その中のスポーツに入ってサッカーを探す。所が、番組数の増大にともなって、メニュー形式での選択が非常に煩雑になってきている。そこで、図6のデジタルテレビでは、音声リモコンを利用して見たい番組を一発で探せる工夫がされている。ユーザーは音声リモコンを手に持って、リモコンのボタンを押すとキャラクターが画面に登場する。キャラクターに向かってサッカーが見たい場合には「サッカー」と発声すると、一気にメニューのサッカーのところまでスキップする。このような対話形式のインターフェースはコンテンツが益々増大する中、今後情報の利用・発信を円滑にする為の鍵になると考えられる。

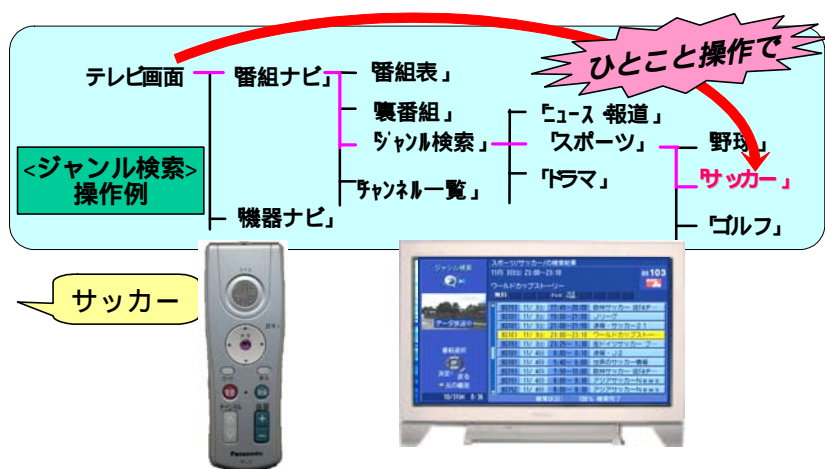


図6 音声認識リモコンによるデジタルTV番組視聴

(3) 基盤技術

多言語音声翻訳[17] [18] [23] [25]

文化芸術という観点から、言語を超えたコミュニケーションは、今後非常に重要になってくると考えられる。海外旅行等で道に迷った場合や、現地出レストランを探す場合

⁵ デジタルデバイス：地理的な制約、年齢・身体的な条件等に起因する情報通信技術の利用機会及び活用能力の格差。

等、とっさには外国語はなかなか出てこないものである。

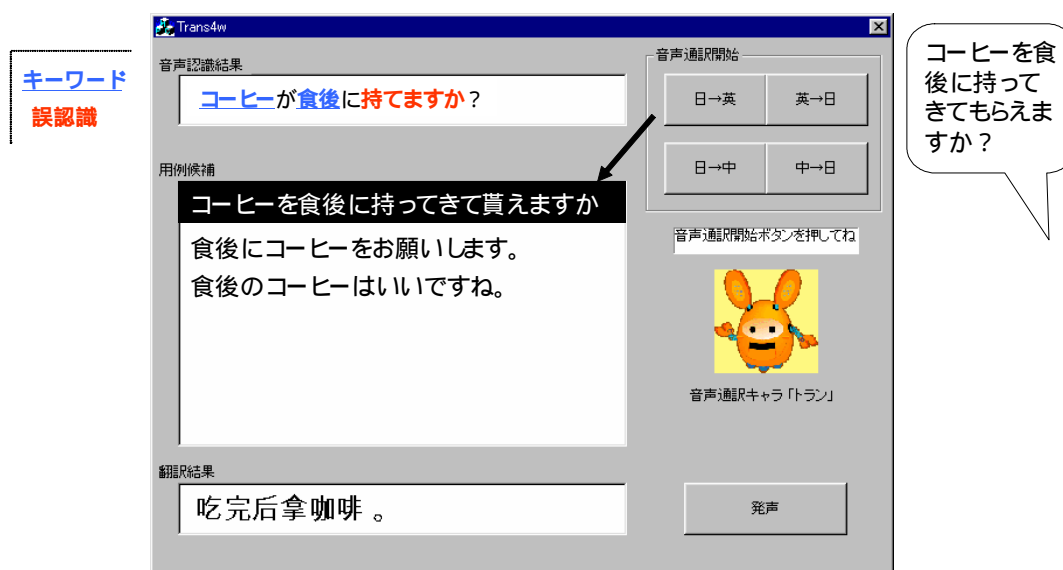


図7 多言語音声翻訳のユーザーインターフェース例

英語であれば、話せる人の数も多くなってきたが、例えば、中国語では不得意な人の割合もまだまだ多く、外国人との会話を支援する翻訳装置に対するニーズは益々増加してきている。そこで、音声認識技術と翻訳技術の単純な組み合わせによる方法が考えられる訳だが、音声リモコンと同様、音声認識率を100%にするのは至難の業である。その理由は、簡単に想像がつく通り、人間同士が会話する場合でも人間の音声認識率が100%になることは非常に稀であり、それでも会話が成り立つ理由は、人間が理解できなかった場合には「何ですか」などと聞き返して間違いを訂正、補完できる点である。これと同じ原理を応用した事例を図7の翻訳装置に示す。

例えば、「コーヒーを食後に持ってきてもらえますか」と日本語で話したとする。発話した文章は、必ずしも正確には認識できないので、「コーヒーを食後に持ってきてもらえますか」という音声データに対して、非常に似た複数の文章の候補から、意味が通じる文章を全部列挙する。つまり、機械の側からユーザーの発話を聞き返して確認するのである。機械が聞き返した文章の中から、実際の意図に合った文章を選ぶことにより、機械からターゲットとなる外国語の音声再生される。一見すると非常にまどろっこしいように感じるが、実際の使用実験では、意外と使いやすいとの結果が得られており、このような対話システムが、今後のインターフェースの大きな鍵になると考えられる。

認知発現発達研究[26] [27] [28]

人間の知能やその仕組みが、どの様に形成されるのかが分かれば、人の知識、匠の技などの暗黙知の形式知化も夢ではなくなり、芸術文化の支援という観点でも非常に役に立つと考えられる。ここでは、人間の物まねを工学的に実現する仕組みの基礎的な研究について紹介する。

この研究は、いわゆる認知工学的な賢い機械、インターフェースを実現することを目指しており、匠の技を形式知に変換するアプローチを採っている。人について考えてみると、赤ちゃんの時には、学習する仕組みはあるが知識は何もない。ところが、赤ちゃんは自分の親あるいは周りの環境と相互作用（インタラクション）することにより、いろいろな知識を獲得していく。それと全く同じことを工学的に辿ることで、一体どこまでできるのかが主な研究内容である。図8に示すように、この研究を通じて構築される発達システムは、人とその周りの環境との相互作用（インタラクション）をベースに、人に違和感を与えない高度なインターフェースを提供することを目指している。これは、3者の関係で物事を捉える「トライフェース」というまったく新しい概念である。

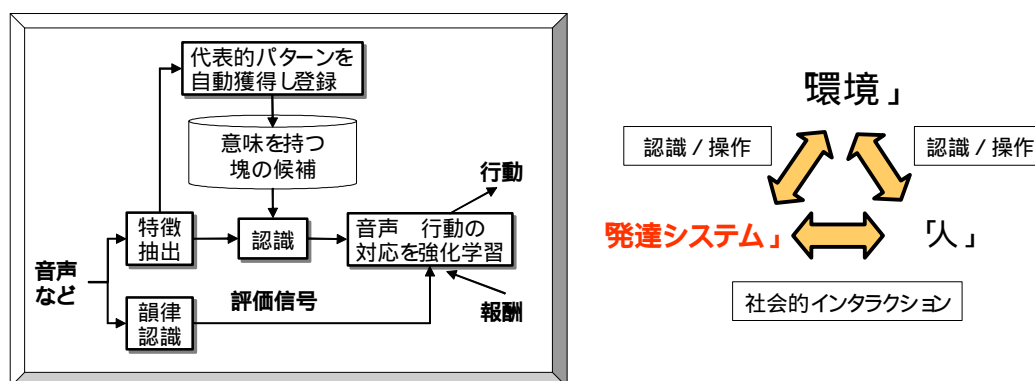


図8 トライフェースの概念

3 - 4 まとめ

本節では、情報技術から見た新たな文化創造と題して、一人ひとりが簡便に芸術文化を創出できる環境を整備し、創造したコンテンツを双方向で意のままに利用するための流通・検索の仕組みを実現する情報技術について事例を挙げながら説明してきた。今後この分野は、万人に安全快適で、自由で豊かな知的創造環境の構築という究極の目標を目指しながら、次々と新しい技術の開発が進み、発展していくと思われる。

過去に遡ってみると、そこには距離や時間の壁があったが、通信技術により、この壁

を突破してきた。現在は、冒頭に説明したようなボタン、リモコンの限界とか、メディアや感覚の壁とか、情報量の壁が立ちはだかってきている。さらに将来的には、コミュニケーションの限界、言語や文化の壁、安全性の壁が待ちかまえている。このような壁を一つひとつ超えていくことができれば、今後も情報技術が日本の芸術文化の発展に深く寄与していくものと考えられる。これからの壁を超えていくには、視聴覚・認知研究、知能・数理研究、人間情報科学研究、あるいは言語モデルや社会情報処理の研究を含めた知能情報技術の確立が極めて重要になってくると考えられる。

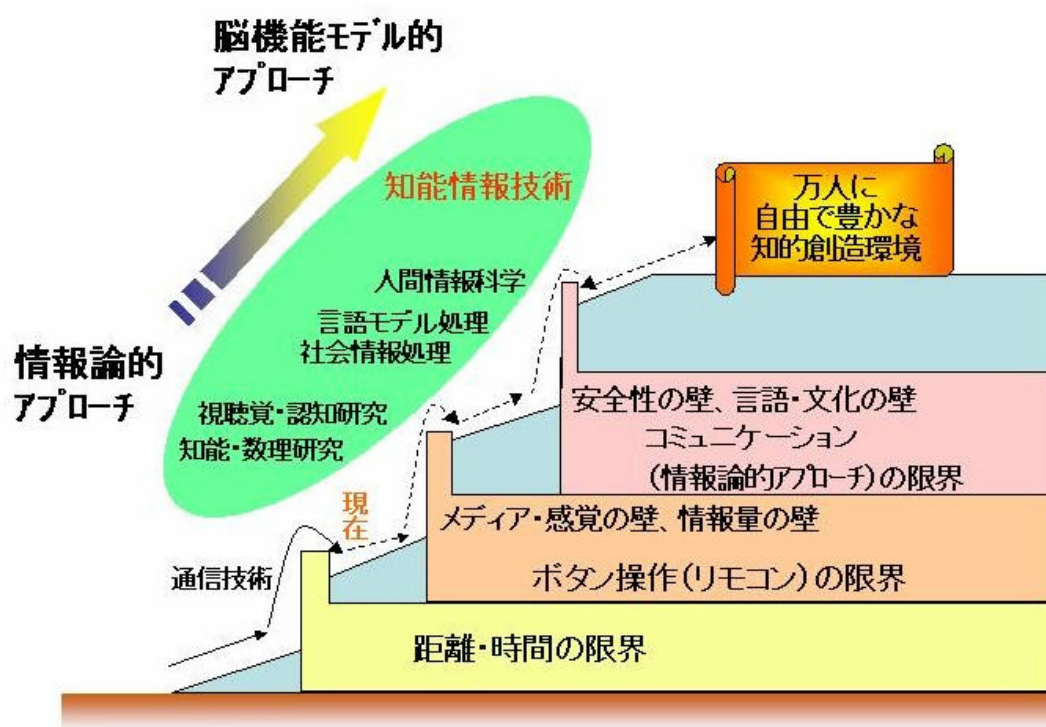


図9 情報技術が目指すべき目標

参考文献

- [1] IEEE LEOS newsletter, 13, 6 (1999) RHK Telecommunications Industry Analysis International Data Corp.
- [2] 「光強度変調によるアクティブレンジファインダの性能評価」映像メディア学会 12,5,1997
- [3] 「マルチスリット方式によるアクティブレンジファインダ」,3,27,1998
- [4] 安方満, 赤井宣子, 山本洋三, 遠藤充: 音声認識を用いたメタデータ制作シ

システム”ボイスタグシステム”,平成15年電気関係学会関西支部連合大会講演論文集, G15-6, 2003

[5]金森、山田、本村、飯川、麓、小寺:「プリズム補間法を用いたカラー画像のRGB-LAB変換」,1992年画像工学コンファレンス,pp.249-252(1992)

[6]小寺、金森「カラー画像工学」(映像情報メディア学会編 日下秀夫監修 オーム社)第5章 pp.109-110,(1997)

[7] 工藤貴弘:概念学習を用いたエージェントによるコミュニティ形成,平成12年度電気関係学会関西支部連合大会講演論文集,pp.G303,2000.

[8]久保田秀和:POC caster:インターネットコミュニティのための会話表現を用いた情報提供エージェント,人工知能学会誌,vol.17,No.3,pp.313-321,2002

[9]西田豊明:コミュニティの知識創造を支援するインタラクティブなメディアをめざして、情報処理,Vol. 41, No. 5, pp. 542-546, 2000.

[10]Toyoaki Nishida: Social Intelligence Design for Knowledge Creating Communities, Invited Paper, The 2001 International Conference on Intelligent Agent Technology (IAT-2001), Maebashi, Japan, October 23-26, 2001

[11] 菊池 義浩,堀 修:”ストリームメディア通信サービス”,情報処理学会 学会誌 Vol.42 No.12 通巻442号pp.1216-1220 2001年 12月

[12]関口,栄藤,江村,藤川,益満,越後:MPEG-7マルチメディア表現に基づくビデオダイジェスト視聴およびオーサリングシステム 信学技報 Vol. 101, No. 302,PRMU2001-92, pp. 51-58 2001.09

[13]井上剛,小沼知浩,西崎誠:”応答文の切り替えがもたらすタスク達成時間短縮効果についての検討”音響学会講演論文集,3-P-25,2002-3

[14]井上剛,西崎誠,小沼知浩,桑野裕康,脇田由実:”EPG番組情報検索向け音声インタフェースにおける認識語彙選択手法の検討”情報処理学会第65回全国大会,2F-1,2003-3

[15]H. Kuwano, K. Fujita, T. Tsuzuki, T. Ishihara:"A Speech Recognition System for Digital TV employing Speaker Adaptation",WESPAC8,2003

[16]K. Fujita, H. Kuwano, T. Tsuzuki, Y. Ono, T. Ishihara:"A NEW DIGITAL TV INTERFACE EMPLOYING SPEECH RECOGNITION",IEEE Trans. On Consumer Electronics Aug, 2003

[17]Y.Wakita, K.Matsui, Y.Sagisaka:"Fine keyword clustering using a thesaurus and example sentences for speech translation",ICSLP2000

[18]C.Zong, Y.Wakita, B.Xu, K.Matsui, Z.Chen:"JAPANESE-TO-CHINESE SPOKEN LANGUAGE TRANSLATION BASED ON THE SIMPLE EXPRESSION", ICSLP2000

- [19]Y.Wakita, K.Matsui, Y.Sagisaka:"Robust speech translation using fine keyword clustering", Workshop on Multi-Lingual Speech Communication 2000
- [20]K.Matsui, Y.Wakita, T.Konuma, K.Mizutani, M.Endo, M.Murata:"AN EXPERIMENTAL MULTILINGUAL SPEECH TRANSLATION SYSTEM", Workshops on Perceptual/Perceptive User Interface 2001
- [21]T.Konuma, K.Matsui, Y.Wakita, K.Mizutani, M.Endo, M.Murata:"An experimental multilingual bi-directional speech translation system", 9th International Conference on Theoretical and Methodological Issues in Machine Translation, 2002
- [22]K.Mizutani, T.Konuma, M.Endo, T.Nambu, Y.Wakita:"Evaluation of a Speech Translation System for Travel Conversation Installed in PDA", IEEE Consumer Communications and Networking Conference, 2004(to be published)
- [23]廣瀬良文,脇田由実,小沼知浩,西崎誠,遠藤充:音声翻訳のための口バスターな類似用例選択手法の評価, 情報処理学会研究報告 SLP-36, 2001
- [24]堀一成,青野繁治,藤家洋昭,石島悌,脇田由実,高階美行:「多言語同時処理」研究の射程と言語間バリアフリー, 情報処理学会第65会全国大会, No.5, pp347-350, 2003
- [25]水谷研治,小沼知浩,遠藤充,南部太郎,脇田由実: PDAで動作する旅行会話向け音声翻訳システムのインタフェース評価, 情報処理学会研究報告 HI-103, 2003
- [26] Koji Morikawa, Natsuki Oka, "Triface: The Interaction between Human, Robot and Environment", Proceedings of the 2003 IEEE International Workshop on Robot and Human Interactive Communication, pp. 187-192, (2003)
- [27] 小松孝徳,鈴木健太郎,植田一博,開一夫,岡夏樹:"パラ言語情報を利用した相互適応的な意味獲得プロセスの実験的分析", 認知科学, Vol.10, No.1, pp.121-138 (2003)
- [28] Yu Ohigashi, Takashi Omori, Koji Morikawa, Natsuki Oka:"Acceleration of Game Learning with Prediction-based Reinforcement Learning - Toward the emergence of planning behavior -", 10th International Conference on Neural Information Processing, pp.786-793 (2003)