

障害者の利用を考慮した音声認識による情報家電の制御 ～ホームネットワークにおけるマルチモーダルインタフェースの提案～

伊藤 英一 神奈川県総合リハビリテーションセンター主任研究員
神奈川リハビリテーション病院
研究部 主任研究員

1. はじめに

重度身体障害者の日常生活を支援する道具のひとつに環境制御装置（Environmental Control System；以下、ECSとする。）がある。利用者の身体機能に応じたスイッチやセンサなどを操作することにより、呼び鈴や電動ベッド、テレビ、ラジオ、電話、エアコンなど日常生活に必要な家庭電化製品（以下、家電とする。）の集中制御をおこなうものである。高位頸髄損傷による四肢まひ者（以下、頸損者とする。）により利用されることが多く、呼気スイッチや音声認識を用いたものが主流となっている。また、操作対象である家電の種類や高機能化により制御指令が増加したため、操作方法に複雑な階層構造を取るシステムが増えている。最近では、複合様式のインタフェース（以下、マルチモーダルインタフェースとする。）が解りやすい操作方法として研究されている。そこで、頸損者の家電操作においてマルチモーダルインタフェースを応用したECSを試作した。これは、音声と動作の複合するコマンドにより、自然で解りやすい操作方法を提供するものであり、その概要と試用評価実験について報告する。

2. 背景

重度身体障害者、特に首から下の身体機能にまひのある頸損者向けに開発されたECSは、日常生活において利用する様々な装置の操作を僅かな身体機能でも利用可能とした¹⁻³⁾。公的助成制度がほとんど無いため、ECSの導入事例は多くないが、導入支援や操作方法に関する要望は多い⁴⁾。最近では音声認識を利用したECSがいくつか製品化されている。

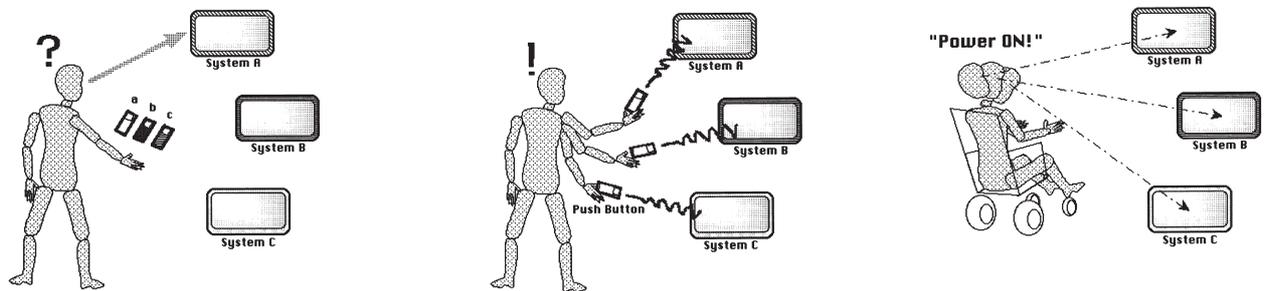
しかし、音声認識の福祉応用への要望は高いが、現状での音声認識のみによる機器操作に関しては操作感の不足など問題の指摘も多い⁵⁾。さらに家電自体が多様化しており、現状のECSの機能だけではニーズに合わないが増えてきている⁶⁾。このような家電の高機能化により、各家電に付属されるリモコンにはボタン数が多くなり、操作方法も複雑で、一般利用者にとっても操作性が良いとは言えない。これら高機能化された家電の制御をECSによって効率よく選択させるためには階層構造を利用しなければならない⁷⁾。しかし、様々な操作様式の家電に対して、安易に階層構造などの複雑な操作方法を提供しては誤操作が多くなり、危険を伴う場面も生じてくると思われる。

そこで、情報機器の使い易さに注目して研究され始めたマルチモーダルインタフェースを応用すれば良い操作環境が提供できるはずである。特に、情報家電の普及に伴い多種多様な操作対象が存在するようになると、操作対象に固有の名称を割り付ける（言語的なタグ付け）よりも、操作対象そのものを非言語的に指示するモダリティが有効である。

複数の家電を操作する環境には複数のリモコンが存在し、どれが目的の家電に対応しているのかで混乱することが多い（図1 a 参照）。複数の家電をひとつのリモコンで操作できるマルチリモコンの場合も同様、目的の家電とマルチリモコンに設定された指令との対応がとれなければ混乱する。そこに、非言語的な選択手段である指差しと、例えば『電源ボタンを押す』という共通化された行為の複合操作を導入することにより混乱は解消されると考えられる（図1 b 参照）。

コンピュータインタラクションにおける非言語的な指差しの研究には、仮想オブジェクトの移動に利用した「Put-That-There」⁸⁾、ネットワークを用いた遠隔共同作業に利用した「ClearBoard」⁹⁾や「超鏡」¹⁰⁾がある。障害者支援関連では、頭部の向きによるコミュニケーション機器¹¹⁾や、頭部の向きによるポインティングデバイス¹²⁾があるが、利用者の残存機能を利用しているにすぎない。視覚障害者用歩行支援システム「トーキングサイン」¹³⁾は指向性を有する信号を受信用ハンドセットで捕らえ、その方向に何かあるのかを伝えるシステムであり、意図的に何かを指示するものとは異なる。

我々は、頭の向きから操作対象を選択する指差し行為と、音声言語による制御指令との複合したマルチモーダルインタフェースが障害者にも自然で解りやすく、今後さらに増加するであろう家電の制御に適した操作システムとして提案する（図1 c 参照）。



a. 複数のリモコン（現状）

b. 方向指示による選択方法

c. 頭の動きと音声で指示

図1.複数の家電におけるリモコン操作

3. 研究の目的

頸損者が多様化・高機能化する家電を自然で理解しやすい手段により利用できるシステムとして、マルチモーダルインタフェースの応用が効果的であると仮定し、その試作と実証実験をおこなう事が本研究の目的である。頸損者には首から下の四肢・体幹に運動まひと感覚まひがある。彼らに適し、かつ自然で理解しやすい手段のインタフェースとしては頭部や眼球の運動、音声などが考えられる。そこで、頭部の運動（非言語的な方向指示のモダリティ）を家電の選択に、音声言語（言語的なコマンド）を家電の制御指令として利用する。このような「頭の向き」と「音声」を利用したマルチモーダルインタフェースによるECSを試作し、複数の家電を操作する試用評価実験によりその有効性を確認する。

4. システムの概要

4.1 システム要件

研究の目的から考えられるシステムの基本要件は、入力操作としての「頭の向き」と「音声」、および結果としての「家電制御」である。さらに、機器制御の観点から逐次変化するシステムの状態を利用者へ伝達する仕組みは不可欠である。伝達する方法は様々存在するが、利用者である頸損者がシステムを操作する状況を予測してみると、視覚は操作対象に向けており、触覚は伝達できる情報量に制限があり、また利用者には感覚まひがあることから、システムの状態を伝達する手段としては聴覚によるものが適すると考えられる。

「頭の向き」を検出する方法としては、複数のビデオカメラによるステレオ画像から算出するシステムや、超音波式マウスエミュレータ（頭の動きでマウスカーソルを移動させるもの）、地磁気から方向を算出するもの、限られた空間に磁界を作りその空間における位置と方向を検出するものなどがある。ただし、利用者は様々な場所・姿勢から操作するため、「頭の向き」とは単なる方向だけではなく、利用者の空間位置をも検出しなければならない。

ステレオ画像によるものは非接触（無拘束）だが高価である。超音波式マウスエミュレータは安価だが、マウスとしての二次元操作しか対応していない。地磁気を利用するものは空間位置の検出が不可能である。

以上より「頭の向き」を検出する方法としては、操作範囲が半径約1.5mと限定され、かつ接触式（ケーブルが接続）であるため操作範囲が制限されるが、6自由度の検出が可能なポヒマスシステム(ISOTRAK II)を利用する。

「音声」による入力操作としては、日本語音声認識システムがある。本システムにおける「音声」とは単なるコマンド入力だが、現在市販されている日本語音声認識ソフトはディクテーション（口述筆記）入力を中心である。そのため、応答性が犠牲となるが入手の容易さから、ピアボイスプロを利用する。

「家電制御」には、iLINK(IEEE1394), HAVi, HomePNA, JINI, ECHONETなどのホームLANシステムが理想的である。しかし、現時点では対応する家電の入手が困難、かつ高価であることから学習型赤外線式マルチリモコン(CROSSAM2+)を利用する。そのため、操作対象には赤外線式リモコンにて制御でき、かつ同じリモコン信号ではない家電を用いるものとする。

聴覚による「状態フィードバック」としては、コンピュータに内蔵されているサウンド出力機能を利用し、WAVファイルをヘッドホンに出力する。

また、これら構成要素を結びつけるための処理ソフトをVisualBasic(VB)で開発し、制御処理用コンピュータ(Pentium III/Windows98SE)で統合処理する。

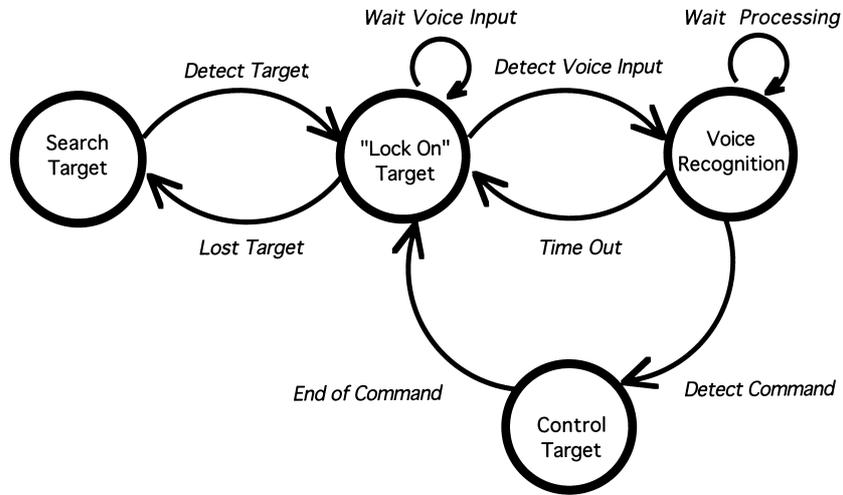


図2.システムの処理フロー

4.2 システムの処理フロー

本システムにおける処理の流れを図2に示す。

- 1) 位置角度センサの情報を一定時間間隔(100ms)にてシステムに取りこむ。
- 2) 三次元位置情報 (X、Y、Z) および三次元角度情報 (オイラー角 ; θ 、 ϕ 、 ψ) から操作空間における操作者の「頭の向き」となるベクトルを演算する。
- 3) 操作対象 (以下、ターゲットとする。) の空間位置と 2) のベクトルを比較し、ベクトルがいずれかのターゲットを貫通していればそれを指していると判断する。
- 4) ターゲットと判断されなければ、引き続き位置角度センサからの情報を待つ。
- 5) ターゲットを指していれば、その状態 (ロックオン状態) をサウンド出力により操作者へフィードバックし、音声入力があるまで待つ。
- 6) ターゲットを見失ってしまえば、その状態 (ターゲット喪失) をサウンド出力で操作者へフィードバックし、1) の初期状態 (ターゲットの搜索) へ戻る。
- 7) 音声認識し、結果がターゲットの操作コマンドであればターゲット制御を実行する。
- 8) 時間オーバー、もしくは認識エラーとなれば 5) の音声入力待ちへ戻る。
- 9) 操作コマンドに対応する制御指令をリモコンに送出し、その状態 (コマンド送信) をサウンド出力により操作者へフィードバックし、5) の音声入力待ちへ戻る。

システム要件、および処理フローから得られるシステム構成を図3に示す。

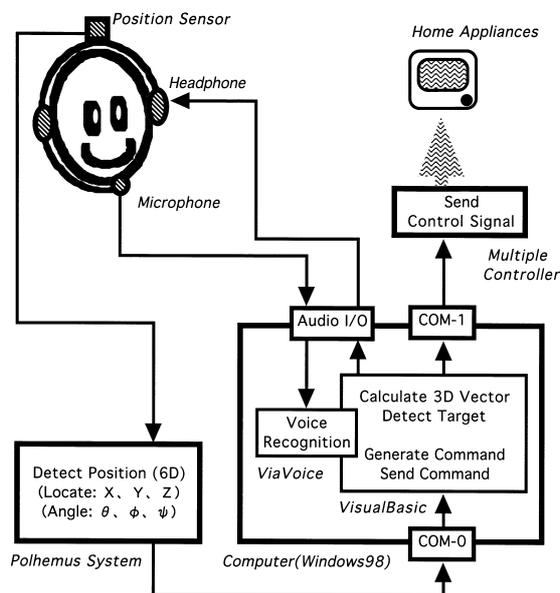


図3.システム構成図

5．評価実験

5.1 操作環境

ポヒマスシステムの利用範囲は半径約150cmである。磁気を利用するため金属（磁性体材料など）などの影響を受けるため、ターゲットの位置や周囲の環境に配慮しなくてはならない。位置角度センサはマイク（ヘッドセット）に装着し、またケーブル類は束ねて衣服に固定し、頭の動きを制限しないように配慮する。

制御処理用コンピュータとポヒマスシステムはラックへ搭載し、初期設定処理やシステムの状態が把握できるよう利用者近傍に設置する。ターゲットは適宜利用しやすい位置に設置し、各ターゲットは指示しやすい間隔（30cm以上）を空けて離す。また、音声認識ソフトの多くは認識率を高めるため、個人の特徴検出のための前処理（以下、エンロールとする。）を必要とする。しかし、本システムでの利用はコマンド入力であることから、特にエンロールは実施しない。



図4.試用評価実験（頸損者）

5.2 試用評価実験

試用評価実験の被験者は、頸損者1名（機能レベルC4：40歳男性）と健常者2名（41歳男性、38歳男性）である。操作姿勢は電動車いす、もしくはキャスター付事務用いすでの椅座位とし、ヘッドセットを装着する（図4参照）。操作姿勢や操作位置はポヒマスシステムの動作範囲内としただけで、特に制限は設けず自由とした。ターゲットは、テレビ、ビデオ、CDラジカセ、電気スタンド（赤外線式リモコンにて操作可）の4種類である。ターゲットへの制御指令としての音声コマンドは事前に利用者へ提示し記憶させておく。音声コマンドは、『電源』：電源スイッチのon/off、『チャンネル』：チャンネルUp、『音量アップ』：音量の増加、『音量ダウン』：音量の減少、『+（プラス）』：増加、『-（マイナス）』：減少、『点ける』：点灯、『消す』：消灯、『プレイ』：再生、『ストップ』：停止、『戻す』：巻き戻し、『送る』：早送り、の12種類とした。音声コマンドとターゲットの組み合わせによる具体的な制御指令を表1に示す。

被験者はこの環境において、自由にターゲットを操作する。操作内容は誤操作評価のためにビデオ収録し、また実験後には制御感やストレス、疲労感など主観的な情報を聞き取り調査する。

表1.各ターゲットにおける制御指令の対応

ターゲット 音声コマンド	テレビ	ビデオ	CD ラジカセ	電気スタンド
『電源』	電源ボタン	電源ボタン	電源ボタン	点灯
『チャンネル』	チャンネル Up	—	選曲 Up	—
『音量アップ』	音量増加	—	音量増加	—
『音量ダウン』	音量減少	—	音量減少	—
『プラス』	音量増加	—	音量増加	明るく
『マイナス』	音量減少	—	音量減少	減光
『点ける』	電源ボタン	電源ボタン	電源ボタン	点灯
『消す』	—	—	—	消灯
『プレイ』	—	PLAY ボタン	PLAY ボタン	—
『ストップ』	—	STOP ボタン	STOP ボタン	—
『戻す』	—	巻戻し	巻戻し	—
『送る』	—	早送り	早送り	—

6. 結果

6.1 「音声」について

数回の試用による操作訓練を実施した後、2回の試用評価実験をおこなった。頸損者には肺活量の低下があるため、マイクの入力ゲインを調整した。特に、認識率に関してはエンロールの有無に関係なく問題とはならなかった。

音声コマンドは12種類と比較的少なく設定したが、錯覚による未設定音声コマンドの入力があった。未設定音声コマンドを入力しても動作しない(未処理)ため、正しい音声コマンドを再入力した。未処理となった未設定音声コマンドは「オン」「オフ」「アップ」「ダウン」「停止」「止まる」の6種類であった。このうち、「オン」「オフ」は反対語であるが、音響的には類似している。つまり、このようなコマンドを設定した場合、誤認識すると動作が反対となる可能性が生じる。このため、音声コマンドとして設定できるものと未設定のままが良いものとの検討するべきである。

6.2 「頭の向き」について

位置角度センサは軽量であるため、ヘッドセット頭頂部に固定しても操作性において大きな問題は生じなかった。ただし、頸損者では電動車いすのヘッドレストにヘッドセットが触れることで、ヘッドセットの装着位置がずれてしまうことがあった。頸損者自身ではヘッドセットの位置を修正できないため、ヘッドセット以外の固定方法も検討の必要がある。

ターゲットを選択するためには頭をターゲットに向ける必要がある。ターゲットをロックオンした時、およびロックオンから外れた時、それぞれのサウンド出力によって利用者へフィードバックされる。しかし、実空間において現在のベクトルがどこを指しているのかが解らないため、どちらへ移動させたらロックオン状態になるのかなどの情報が欠ける。そのため、ロックオン状態にする、もしくはロックオン状態を継続しながら音声コマンドを発声するためには、体性感覚のみによる頭の姿勢制御であり容易とは言い難い。

6.3 システムについて

頸損者は起立性低血圧などへの配慮から電動車いすや電動ベッドの背もたれを定期的に倒すことが多い。背角度が倒れている状態では、例えばターゲットが視野に入っていたとしても頭部は上向きでありターゲットを指示す事が困難となる場面があった。実験後の主観的な聞き取り調査では、既存ECSよりも操作が容易との意見以外には、システムに対する質問が多く明記すべき情報は得られなかった。

以上、特殊な場面における問題点は見つけられるものの、設置した全てのターゲットを自由に使いこなす事が確認で

きた。また、音声コマンドを共通化させ、さらにターゲットを選択するための言語的なコマンドをなくすことでコマンド数を減少できたことから、僅かな操作説明だけで利用可能となった。つまり、操作方法が解りやすくなったのではないかと考えられる。

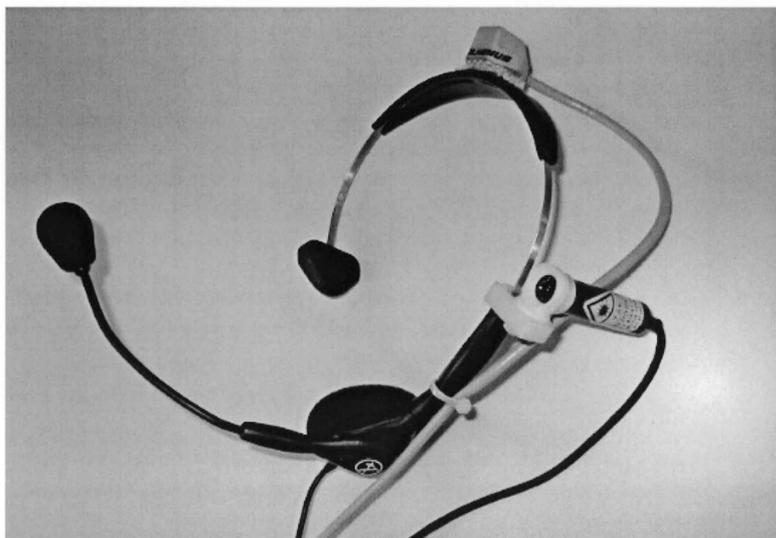
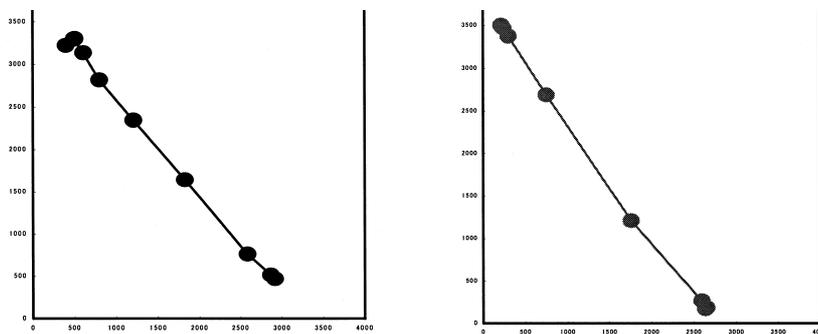


図5.レーザーポインタ付きヘッドセット

7. 考察と今後の課題

まず、操作性を一層高める方法として考えられることは、実空間における頭の向き（ベクトル）を視覚的にも表現し、利用者へフィードバックさせる事である。実世界空間の特定対象物を指示するものとして「GestureCam」¹⁴⁾がある。これは遠隔コミュニケーションにおいて、相手側の実空間をレーザーポインタで指示するシステムである。そこで、レーザーポインタを利用し、位置角度センサのベクトルを視覚化するためのヘッドセットを試作した。（図5参照）

このレーザーポインタの付加によって、ロックオン状態にする、もしくはロックオン状態を維持させるための姿勢制御を体性感覚に追加して視覚的フィードバックも併用することから操作性は高まると考えられる。そこで、簡単な実験として、ターゲットをロックオンさせるまでの軌跡を、レーザーポインタの有無により調査した。その結果、レーザーポインタを装着した場合には装着しなかった場合と比較して、移動開始からロックオンさせるまでの時間が短く、さらにターゲット近傍での位置合わせ動作が減少していることが認められた（図6参照）。ただし、実際の利用においては危険なレーザー光が周囲の人間の眼に入らないようにする工夫（例えば、頭の向きに応じて一定時間発光する）が必要であろう。



レーザーポインタ無し

レーザーポインタ有り

図6.レーザーポインタの有無によるロックオンまでの軌跡の変化

電動車いすや電動ベッドにおける利用では、背角度などの姿勢に応じてヘッドセットのベクトルに上下方向のオフセット角度を加算するような仕組みが有効だと考えられる。また、位置角度センサとして利用したポヒマスシステムは利用範囲が狭く有線式であり、一般的なインタフェースとしては高価である。情報家電への対応に向けたホームLANへの拡張も必要であろう。

本システムの一般利用者向けへの応用例として、「音声」コマンドの代わりに押しボタン操作を、また「頭の向き」の代わりに押しボタンの向きとするリモコンを試作した(図7、図1b参照)。このリモコンでの操作では、視線との差(視差)がヘッドセットの時よりも大きくなるため、一層レーザーポインタの必要性が感じられた。つまり、実空間における実オブジェクトへのポインティング操作を考えると、空間ベクトルの可視化が大きな課題である。



図7.方向指示によるボタン型リモコン

まとめ

「音声」と「頭の向き」によるマルチモーダルインタフェースを用いた、頸損者が解りやすい家電操作システムを試作した。試用評価実験から、提案した操作方法が頸損者に有効であることを明らかにした。これは音声コマンドの共通化によりコマンド数を絞り込めた事と、ターゲット選択の指示を非言語化した事により、覚えるべき操作を減らすことができたことによると考えられる。また、制御したいターゲットに頭(顔)を向けることで、機器選択が可能となるため、制御すべきターゲットが増えたとしても、音声指令は増えない。このように、「音声指令(言語的)」と「頭の向き(非言語的)」のマルチモーダルインタフェースは、障害のある利用者にも有効であることが明らかとなった。特に今後普及するであろう情報家電のコントローラとして応用されていくことが望まれる。

参考文献

- 1) 市川冽他(1985), 環境制御装置設置マニュアル, 環境制御装置研究開発連絡協議会編
- 2) 伊藤英一他(1992), 環境制御装置の工夫例, 第7回リハ工学カンファレンス講演論文集, pp.83-86
- 3) 畠山卓朗(2000), テクノロジーの発展と障害者の生活 - 電子機器を中心に, 障害者問題研究, Vol.27, No.4
- 4) 横田恒一他(1997), 環境制御装置に関するアンケート調査, 第12回リハ工学カンファレンス講演論文集, pp.463-466
- 5) 太田将主他(2000), 障害者のための音声インタフェースの一考察, 第15回リハ工学カンファレンス講演論文集, pp.327-330
- 6) 山岸秀之他(2000), 頸髄損傷による四肢まひ者の生活と環境制御装置の課題, 第15回リハ工学カンファレンス講演

論文集, pp.605-608

- 7) 吉田直樹(1997), 選択メニュー学習機能付き環境制御装置の提案と操作効率シミュレーション, 第12回リハ工学カンファレンス講演論文集, pp.453-458
- 8) Bolt,R.A.(1980), Put-That-There: Voice and Gesture at the Graphics Interface, ACM SIGGRAPH ' 80 Computer Graphics, 14(3), pp.262-270
- 9) Ishii,H., et al.(1993), Integration of Interpersonal Space and Shared Workspace: ClearBoard Design and Experiments, ACM Trans. Information Systems, Vol.11, No.4, pp.349-375
- 10) 森川治他(2000), 人に優しい超鏡対話における指差し行為, 情報処理学会論文誌, Vol.41, No.5, pp.1290-1297
- 11) 高見修他(1997), 頭部の向き検出と介助機器への応用, 第12回リハ工学カンファレンス講演論文集, pp.183-186
- 12) 小泉義樹他(1997), 高位頸髄損傷者用パソコン入力装置の開発, 第12回リハ工学カンファレンス講演論文集, pp.97-102
- 13) 畠山卓朗他(1998), 音声歩行案内システム, 第14回ヒューマンインタフェースシンポジウム前刷集, pp.577-583
- 14) Kuzuoka,H., et al.(1994), GetrureCam: A Video Communication System for Sympathetic Remote Collaboration, Proc. CSCS ' 94, pp.35-43

< 発 表 資 料 >

題 名	掲載誌・学会名等	発表年月
Multi-modal Interface with Voice and Head Tracking for Multiple Home Appliances	In proceedings of the 8 th IFIP Conference on Human-Computer Interaction (INTERCT 2001), IOS Press	2001年7月
頸髄損傷者のためのマルチモーダルインタフェース～使いやすさに着目した家電製品の制御方法～	第16回リハ工学カンファレンス (日本リハビリテーション工学協会)	2001年8月