

音声言語を用いた仮想空間との対話による試着システム

河原 達也[†] 田中 克明[†] 堂下 修司[†]

仮想空間技術と音声対話技術を融合したヒューマンインタフェースを提案する。仮想空間技術は、実空間と同様の行為で、実空間では実現できないようなことをも可能にする。音声対話技術は、音声言語を介した概念的・感性的な指示を対話的に行うことにより、ユーザの意図を実現する。本研究では、仮想空間との対話における音声言語の特徴を分析し、音声言語が効果的に作用するタスクとして、仮想試着システムを設計・作成した。本システムは、ユーザが試着室にいるような状況を想定し、音声言語による色や柄の指示に基づいて、仮想の鏡に試着イメージを合成する過程を対話的に行うものである。音声言語による直接的な指定の結果を対話的に画像合成することにより、ユーザが主導権を持ちながら、自然でかつ頑健なインタラクションを行えることが確認できた。

Virtual Fitting Room with Spoken Dialogue Interface

TATSUYA KAWAHARA,[†] KATSUAKI TANAKA[†] and SHUJI DOSHITA[†]

We present a novel human interface that incorporates spoken dialogue into virtual space technologies. The virtual space enables us to do what cannot be done in real world as if we do in real world. Spoken dialogue technologies realize our intention by understanding specifications that include concept and a nuance. We clarify the properties of the speech interface with the virtual space, then design and implement the virtual fitting room, where the speech interface works effectively combined with virtual space technologies. The system supposes that a user is in a fitting room. The user can select his/her favorite clothes interactively by telling his/her preferences and checking with the virtual mirror. It is observed that the combination of spoken language input and virtually real image output realizes natural and robust interaction while the user keeps the initiative.

1. はじめに

近年の計算機技術および通信技術の進歩にともない、遠隔会議システムや遠隔ロボット操作システム等の仮想空間技術を利用したシステムが数多く実現されている¹⁾。この仮想空間との対話手段として現在主に用いられているのはマウスやデータグローブといった視覚を利用したポインティングデバイスである。視覚を利用するため、これらの入力手段には操作が直感的で分かりやすいといった利点があるが、その一方で操作対象が画面内に映っているものに限られたり、もどかしさを感じる場面も多い。

日常生活における人間の活動では言葉を用いてやりとりがなされることが多いことから、仮想空間とのインタフェースの手段としても音声言語を用いることは有効であると考えられる。特に近年、音声対話に関する研究がさかに行われ、成果を収めつつある。しか

し従来、仮想空間との対話において音声は用いられないか、用いられても“put that there.”²⁾のように補助的に用いられることが多い³⁾。

仮想空間を直接的に対象としていないが、近年マウス等のポインティングデバイスと音声とを統合したマルチモーダルインタフェース^{4),5)}の研究がさかに行われている。これらの例には、作図ツールのようにポインティングデバイスと協調的に空間的な操作を行うもの^{6),7)}と、地理的な情報を利用しながら固有名詞などの入力を効率的に行うもの^{8),9)}があり、いずれも操作時間の短縮によりその有効性を示している。

ただし、単純な入力作業における長時間の音声入力は疲労をとめない必ずしも実用的ではないとの指摘³⁾もあり、我々は音声言語に特有で他のモダリティでは実現できない機能を実現することが重要であると考える。

そこで本論文ではまず、仮想空間との対話における音声言語の特徴を分析して、(1) 音声言語でないといけない指示・操作、(2) 音声言語を用いることにより円滑になる指示・操作に関する分類を行う。そしてこ

[†] 京都大学大学院工学研究科情報工学教室

Department of Information Science, Kyoto University

の考察に基づいて設計・作成した仮想試着システムについて述べる。

2. 仮想空間との対話における音声言語の役割

まず本研究で想定する仮想空間を定義し、次にその仮想空間を用いることで得られる利点について考察する。そして、仮想空間との対話における音声言語の特徴を明らかにする。

2.1 仮想空間の特徴

本研究では仮想空間を以下のように考える。

“実空間と同様な行為で、実空間では不可能なことをも実現するために合成された空間であり、対話の相手として存在する。”

多くの仮想現実感の研究は、特殊ディスプレイやデータグローブなどを用いることにより、ユーザがその空間内に没入しているような感覚を実現しようとしているが、本研究ではコミュニケーションの相手としての空間としてとらえているため、没入感は追求しない。ただし、ある程度の現実感重要であると考え、対象を通常の計算機のディスプレイにアイコン的に表示するだけでは、仮想空間とはよばない。実物に近い大きさや動きをユーザが感じられることにより、従来と異なった形態のコミュニケーションが実現できると考えられる。

特に、実空間と同様の行為で対話できること、および行為の結果を対話的に実感できることが重要であると考え。これらは、インタフェースを直接的・直感的にして、システムの操作性を向上させる。さらに、インタラクションを活発にかつ自然にする効果も期待できる。

また、仮想空間上では実空間で不可能なことも実現できる。たとえば、膨大な在庫を持つ、遠隔のものを瞬時に運ぶ、時間を戻して操作をやり直す、などといった空間的な制約や時間的な制約を超越した操作が可能になる。これは、ヒューマンインタフェースの改善にとどまらず、実現できるタスクの内容・仕様の拡大につながる。

2.2 仮想空間との対話における音声言語の特徴

本節では、仮想空間との対話における音声言語の特徴を、主に種々のポインティングデバイスと比較しながら考察する。その長所として、音声言語により可能となる指示・操作と、音声言語を用いることで円滑になる指示・操作に分類して考える。なお以下の多くは、音声入力に限らずテキスト入力を含めた言語系全般にいえることであるが、本研究では現実感と操作性を追求するため、言語の入力をテキスト（キーボード）で

なく音声で行う。

2.2.1 音声言語により可能となる指示・操作

音声言語でのみ可能となる指示・操作には以下があげられる。これらの機能を導入することにより、音声言語インタフェースが効果的に作用すると期待される。空間内に明示的に表示されていない対象物へのアクセス 音声では空間内に見えていない対象へのアクセスが可能であるため、仮想的な引出しや倉庫に膨大な量のストックを準備して、容易にアクセスすることができる。

属性の指示 “赤い方...”, “子供用...”といった対象の属性に関する指示が可能であり、このような指示を自由に組み合わせ、より複雑な指示を行うこともできる。メニューなどのGUIのみではこれらの記述が大変煩雑であり、実現できたとしても項目が膨大になる。

ニュアンスによる指示・感性的な指示 “派手なもの”, “もっと明るいもの”といったニュアンスによる指示・感性的な指示が可能であり、これらの指示からユーザの意図を正しく抽出できれば、インタフェースが格段に良くなる。

両手を使えない、または使いたくない状況での指示・操作 音声でのやりとりには両手を用いる必要がなく、何らかの行為を行いながら質問等をするには音声に適している。また両手を拘束しないことは、より自然なマンマシンインタフェースの要素になる。

2.2.2 音声言語により円滑になる指示・操作

音声言語を用いることで円滑になる指示・操作には以下がある。これらの特徴は、音声インタフェースの使用を優位にする可能性がある。

多数の選択肢からの直接的な選択 選択肢の数が多数になる場合、メニュー等では目的のものがどこにあるのか分かりにくい。逆に求める対象物の名前が分かっていたら、それをメニューや空間内から探し出すのは非効率であり、このような場合には音声に適している。

距離的に遠い対象物へのアクセス 音声による指示は対象物までの空間的な距離に関係なく一律であるため、特に遠い対象物にアクセスする際にポインティングデバイスよりも効果的である。たとえば、“左から2番目”とか“奥の方”といった指示が可能になる。ただし、空間位置の細かい指定は困難であるので、これらは他の（色や大きさなどの）属性による指示と組み合わせることにより一層効果的となる。

操作への習熟 特に計算機になじみのない人にとって、

日常的な話し方で指示ができれば、機器に慣れる必要がないので、円滑に使用することができる。

2.2.3 音声言語インタフェースの短所

音声言語インタフェースにも当然短所は存在し、システムを設計する際にはその点を考慮すべきである。ここでは音声の持つ短所のうち、システム全体の性能に対して特に影響が大きいと考えられる3点について、その対策も含めて考察する。

第1に、空間の扱いに適していないことである。これは自然言語が一般的にトポロジーや距離感の記述には適していないためであり、システムの設計に際しては、空間操作に他のモダリティを利用するなどの考慮が必要である。

第2に、認識誤りが不可避なことである。対話システムで認識誤りが生じると、ユーザの望んだ動作とは別の動作が行われる。システムは認識誤りから回復する手段を備えておき、その後の対話によって最終的にユーザの意図が実現できる枠組みを用意する必要がある。

第3に、システムが受理できる語彙や構文をユーザが必ずしも明確に把握できないことである。これに対しても、受理できる部分を検出し、受理できない部分を棄却して、やはり対話の枠組みで確認等を行いながら理解を進める方式が望まれる¹⁰⁾。

2.3 音声対話システムにおける仮想空間の効果

本節では、仮想空間を使用することによる音声対話システムへの効果について考察を行う。

音声対話システムにおいては、頑健性が大きな課題となっている。前述のとおり、音声認識誤りは不可避であるうえに、ユーザは、システムの語彙や構文をすべて記憶しているわけではないので、システムの想定外の発話を行うことがある。システムが扱えない発話についても、多くの場合認識誤りとなる。音声対話のみで認識誤りからの回復を行うためには、システムの認識結果をユーザに確認することが必要となるが、音声による頻繁な確認は煩わしいものである。それに対して、マルチモーダルな応答により、システムの処理結果(状態)をユーザに提示することが有効である⁵⁾。仮想空間上にユーザの指示の処理結果がただちに反映されれば、ユーザはシステムの誤りを察して、言い直したりすることが可能になる。つまり音声言語による明示的な確認を行うことなく、インタラクティブに誤りの検出と訂正ができる。

このインタラクションは、ユーザの試行錯誤を含んだ思考過程をも反映することができる¹¹⁾。ユーザは、最初から自分の希望が明確に言えるとは限らず、い

ろ試してみることにより、納得のいく結果に到達することが多い。特に(通常の計算機ディスプレイでなく)仮想空間を用いることにより、対話がより自然になることが期待できる。インタラクションは、ユーザに対してシステムが優れた反応を行うほど活発になると考えられる。現実に近い画像を実時間で連続的に合成することにより、この効果を実現する。結果として、ユーザの意図の達成度や満足度を高めることが期待できる。

3. 仮想試着システムの設計

上記の考察に基づいて、音声言語が効果的に働くタスクとして服装の選択を設定し、仮想試着システムを設計・作成した。

特に、(1)ポインティングデバイスやタッチパネルなどの手や動作を拘束する入力を行いたくない状況を設定すること、(2)属性やニュアンスによる概念的指示が中心になること、の2点に焦点をおいて、ユーザにとって自然で簡単にインタラクションが行えるシステムを設計した。従来の音声入力を用いたマルチモーダルシステムにおいては、計算機・端末がユーザの相手になるために他のモダリティと競合して、音声を使うことの必然性・優位性が必ずしも明確ではない場合が多くあった。これに対して本研究では、試着室という(他のモダリティのために)両手を拘束するのが不自然な仮想空間を設定し、服装の選択という概念的な指示が中心になるシステムを作成した。

本章ではこのシステムについて説明する。

3.1 システムの概要

通信販売が身近なものになって久しく、また最近ではインターネット上での買い物も珍しくなくなりつつあるが、カタログやネットワーク端末を用いて選択するのが困難な商品も多数存在する。服はその代表例であり、カタログや画面上のサンプル写真ではなく、実寸に近い大きさで見たり、あるいは自分に合わせてみたりしないと納得しない消費者が多い。

本研究では、仮想空間の技術によりこの解決を図る。具体的には、仮想的な試着室を構成し、鏡を見ているような感じで服の選択を行えるようにして、手元に実物がない状態での商品選択を容易にする。仮想店舗では、実空間では不可能な膨大なストックを用意し、瞬時に“着替える”ことが可能である。また服をオーダーメイドする際に、その仕上がりを予想するのにも利用することができる。

そこで、仮想試着システムを設計した。本システムは、計算機上に仮想的なストックを用意し、大スクリー

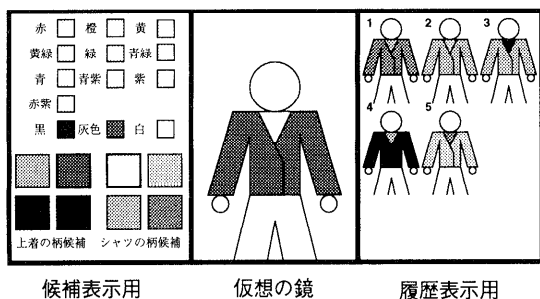


図1 仮想試着システムの画面構成

Fig.1 Display image of virtual fitting room.

鏡と想定して、仮想試着室を構成する。ユーザはこの仮想鏡の前で音声言語を用いて対話的に服装の選択を行うことができる。

本システムにおいてユーザが選択するのは、服の布地である。服の選択の際には、そのデザイン（形）とそれに使用する布地の要素があるが、本タスクでは布地の選択のみとした。選択対象とする布地は、上着とシャツの2種類である。

本システムのスクリーンの様子を図1に示す。ユーザは、音声言語により自分の希望の色や柄を伝える。これを音声認識することにより、それに合った候補をデータベース（ストック）の中から取り出す。特に柄に関しては、音声言語による指示が困難であるので、候補をディスプレイに表示して、ユーザが選択できるようにする（図左）。指定・選択された色や柄に応じた服をユーザの等身大の画像に重ね合わせて、大スクリーンに投影することにより、あたかもその服を試着して鏡に向かってるようにみせる（図中央）。ユーザ画像の入力、および試着イメージの合成は連続的に行うので、ユーザの動きに追従する。ユーザは、自分に合っているか確認しながら、納得するまで音声対話で操作を繰り返す。ある程度気に入った画像は、履歴として保存しておいて、それを参照・比較することが可能である（図右）。

このように本システムは、3枚の画面を使用する。中央の画面が仮想の鏡である。左の画面は柄の候補の表示用であり、右の画面は履歴表示用の画面である。

本研究では、ユーザからの指示の手段として音声言語のみを用いる。これは、試着室にいるような感覚を実現するためには、ポインティングデバイスなどの使用は不自然であることと、システムの使用方法を単純にするためである。

3.2 本システムにおける音声言語の効用

本タスクは、2.2.1項であげた音声言語により可能となる指示・操作をすべて含む。すなわち、明示的に

表示されていない膨大な仮想的なストックに直接的にアクセスすることができる。また、属性やニュアンスによる指示をそのまま自然に行うことができる。さらに、両手を拘束しない状態で、“試着”してそのイメージを確かめることができる。

また、2.2.2項であげた音声言語により円滑になる指示・操作も実現される。特に、色や柄の種類は多数存在するので、音声により効率的に選択することができる。

次に、2.2.3項や2.3節で述べたように、音声言語を使用する際に問題となる認識誤りや想定外発話への頑健性が、仮想空間とのインタラクションにより実現できると考えられる。ユーザは、意図どおりの動作が行われなかった場合でもその状況をただちに察することができるので、再度言い直したり、表現を変えたりすることにより、回復することが可能となる。実社会での店員とのやりとりでも本質的に同様のことを行っていると考えられる。

3.3 システムの構成

本節では仮想試着システムの構成と処理の流れについて説明する。システムのモジュール構成を図2に示す。

ユーザの発話は、マイクروفोनに入力され、音声認識されて、コマンドとして変換される。対話履歴や画面の情報などに基づいてコマンドの省略補完や解釈がなされる。色や柄の検索コマンドであれば、データベースへのアクセスが行われる。解釈およびデータベース検索の結果に基づいて、画像合成部への指示が行われる。具体的には、選択されたテキストチャプターが与えられる。

同時に、ユーザの画像はカメラから入力されており、服の領域の抽出が行われている。服の領域に指定されたテキストチャプターをはめ込むことにより、試着イメージを合成する。画像認識と合成は連続的に実行され、スクリーンに動画として出力される。ユーザの指示が解釈できないなど、画像合成に至らない場合は、音声応答を行う。

次に、各構成要素について説明を行う。

3.3.1 音声認識部

音声認識部は、入力された音声を単語列として認識し、コマンドに変換する。

本システムでは、我々が開発した音声認識パーザを使用した。この音声認識パーザは不特定話者音素モデルにHMMを用い、オートマトンパーザによる構文解析を行う。また探索手法には、単語間の接続に関する制約である単語対文法をヒューリスティックとして用

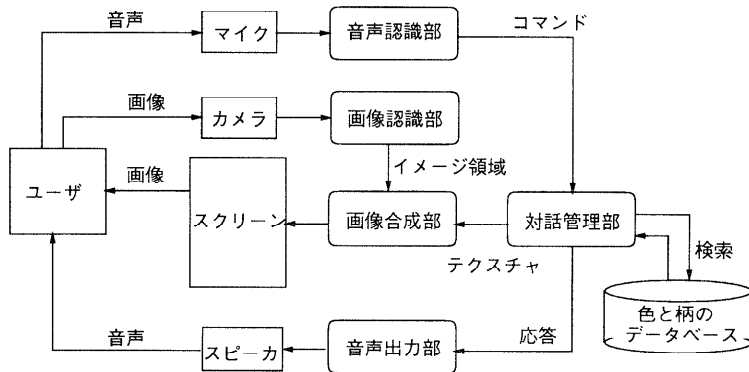


図2 システムのモジュール構成

Fig. 2 System configuration of virtual fitting room.

いたA*探索を採用している¹²⁾。なお本システムにおける発話のパターンは正規文法で記述されており、語彙数は83である。

音声認識結果からタスク遂行に必要なキーワードを抽出して、コマンドへの変換を行う。コマンド形式は、コマンドのタイプ（色名の指定、明度の指定など）、その対象（上着かシャツ）、ユーザの指定した属性（色名など）から構成される。

3.3.2 対話管理部

対話管理部は、コマンドの省略補完・解釈を行い、ユーザへの応答を生成する。

対話の履歴を管理することにより、たとえば、コマンドの対象（上着かシャツ）が省略されたときには、直前のコマンドのものを補完するので、ユーザは発話のたびに対象を指定する必要はない。コマンドの解釈結果に基づいて、指定された色や柄の検索を実行する。また、候補画面や履歴画面が参照された場合は、それら呼び出す。音声応答が必要な場合は、応答文を生成する。

3.3.3 画像認識部

画像認識部は、カメラで入力されるユーザの画像から上着とシャツの領域を分離・抽出する。

衣服の正確な形状抽出には、立体的な視覚情報を利用することが望ましいが¹³⁾、処理時間が大きくなることから、本研究では実時間性を優先して、簡易な方式を採用した。これはクロマキー合成に類似している。すなわち、ユーザにあらかじめ特定の色の服を着てもらい、その色情報のみを手がかりに領域を抽出し、はめ込み合成を行う。ここでは、赤色のシャツと青色の上着を着てもらった設定にした。入力された画像データを順に画素単位でRGB値をHSB値（Hue：色相, Saturation：彩度, Brightness：明度）に変換し、背景や照明などを考慮して設定したしきい値により、上

着とシャツの領域抽出を行う。

3.3.4 画像合成部

画像合成部は、選択されたテクスチャを画像認識部で抽出された服の領域にはめ込むことにより、試着イメージを合成し、実物大でスクリーンに投影する。上着またはシャツと判定された領域には、それぞれに指定されたテクスチャが合成されるが、このとき原画像の明るさの情報（Brightnessの値）を反映させることにより、服の陰影を表現する。

3.3.5 データベース

データベースは柄と配色を別々に管理しており、テクスチャは両者を決定することにより生成される。

柄のデータは、縦じま/横じま/チェックに対してそれぞれ4種類用意されており、いずれも3色から構成される。この3色は1色の主色（メインカラー）と2色の補助色からなっており、主色はユーザの要求に応じて、色相・彩度・明度を変化させることにより決定される。そのため、1,677万色の種類が可能である。補助色は主色の類似色、もしくは反対色のいずれかから選択される。この選択はユーザの選好を反映し、派手な服を好む場合は反対色を、地味な服を望む場合は類似色を選択する。

3.3.6 音声出力部

音声出力部は、対話管理部で生成された応答文を音声として出力する。

3.4 システムの命令体系と語彙

本システムでは服のデザインは扱わないため、ユーザの指定は布地に関するものに限定される。また、布地の質感は扱わず、ユーザには柄と配色に関する指示を行ってもらう。

“花柄”というと赤やピンクの色で配色されることが多いように、柄と配色の間には相互関係があるが、ここでは扱う柄をチェックや縦じまのように幾何学的

表1 コマンドの一覧
Table 1 List of commands.

言語による選択	
● 色名での指定	
例	上着を赤にしてください。
例	上着とシャツを黄色にしてください。
● イメージ語での指定	
例	シャツを明るくしてください。
例	少し地味にしてください。
● 柄名による指定	
例	上着をチェックにしてください。
例	縦じまにしてください。
画面上からの選択	
● 候補画面からの柄の選択	
例	上着を上柄にしてください。
例	下の柄にしてください。
● 履歴画面へのアクセス	
例 (保存)	これを覚えておいてください。
例 (復帰)	3 番に戻ってください。
制御命令	
● 動作のやり直し	
例	やり直してください。
● 終了	
例	終わります。

模様のみとし、柄と配色を分離した。したがって柄に対する命令と配色に対する命令は独立である。

本システムで扱える命令を表1に列挙する。

配色に対する指定は、色名での指定かイメージ語での指定で行う。

柄の選択は、柄名での指定と候補画面からの柄の選択で行う。ユーザが柄名で指定をすると、柄選択画面に複数の候補が表示されるので、ユーザはさらにその中から選択を行うことができる。

動作のやり直し命令は、誤認識が生じた場合の回復手段である。履歴画面へのアクセスは、気に入った服の参照・比較のためのものであるが、誤認識からの回復手段としても用いることができる。

文法はこれらの命令ごとに記述し、ユーザには事前にそのパターンを提示する。この文法は非常に定型的なものであり、ユーザの発話パターンを制限するが、ユーザの理解のしやすさと認識性能を優先させた。より自由な発話からキーワードやフレーズを直接抽出することにより認識を行う方式についても研究を進めており¹⁴⁾、本システムにも今後採用していきたい。

語彙に関して問題となるのが、配色に関するイメージ語の使用である。

色を規定する色相、彩度、明度の3要素は、人間が普遍的にいだく特定のイメージと対応しており、そのイメージを表現する言葉は広く使用されている。具体的には、色相は“冷たい”、“暖かい”、彩度は“派手”、

“地味”、明度は“明るい”、“暗い”といったイメージを表す。このような普遍的なイメージ語は使用できるようにした。

これに対して、“さわやかな”とか“すがすがしく”といったイメージ語は、容易に規定できるものではなく、また感性に深く関係するため、個人差が生じる。そのため指示も曖昧になり、システムの対応を決定することが非常に困難となる。したがって、本研究ではこれらは取り扱わないこととした。

これらの用語と色の対応を調べるには、数多くの心理の実験や専門的な知見が必要である。たとえば、イメージ・スケール^{15),16)}では、平面上に180のイメージ語と配色をそれぞれプロットしており、両者の対応をとることができる。これらの利用は今後の課題である。

4. 実装と考察

本章では、システムの実装について説明し、使用者の評価に基づいて考察を行う。

4.1 実装

本システムは、京都大学工学研究科情報工学教室のヒューマンメディア実験室において、背面投影型プロジェクタと80インチスクリーン各3台を用いて実装された。また、音声入力と画像入力にSGI社のIndyワークステーション各1台、音声認識等にSun Ultraワークステーション、画像出力にSGI社のOnyxを使用した。

音声入力から音声認識・コマンドの解釈までは2~3秒で実現されている。画像入力から画像合成までは1~2秒の遅れで実現されており、毎秒数フレームの処理を行っている。結果として、ユーザの指示から試着イメージの合成までは数秒で実現されている。

図3に柄候補リストの画面例を、図4に仮想の鏡の画面例を示す。

4.2 評価と考察

本システムを10人(20~30代、男性)の被験者に使用してもらった。システムの説明を簡単にした後、自由に服装の選択をしてもらった。ユーザの発話のほとんどは正しくコマンドに解釈された。また、認識誤りが生じた場合でも、言い直したり、動作のやり直しを指示することにより、対話が回復できた。

本システムを、音声言語による指示ではなく色や柄のメニューからマウスで選択するシステムと比較した。メニュー選択画面には、色候補として11色相(図1の候補表示用画面のもの)に関して、明度4段階・彩度3段階の合計132種類を表示し、柄候補としては選ばれた色とパターン(縦じま/横じま/チェック)に対

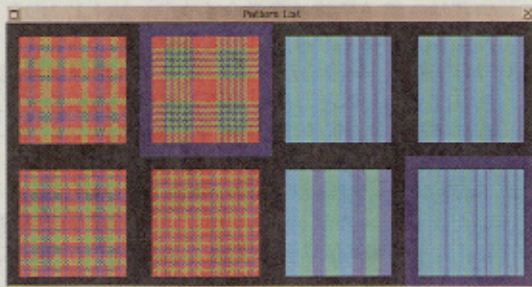


図3 柄候補表示の画面例

Fig. 3 A shot of candidate display.



図4 仮想の鏡の画面例

Fig. 4 A shot of virtual mirror.

して4種類を表示した。10名の被験者に両方のシステムを使用してもらった後で、「選択の容易さ」と「使用後の満足度」の項目に関して、両者の比較を行ってもらった。その集計結果を図5に示す。大半の被験者が、マウスの方が選択が容易であると答えている。マウスによるメニュー選択では、選択画面に出ているものがそのまま試着イメージに合成されるので選択の結果がほぼ確実に予期したものになるためであると考えられる。しかし満足度については、過半数が音声入力をあげている。被験者のコメント等からこの理由としては、概念的な指示・感性的な指示から意図を抽出して、予期していた以上のものが提示されること、メニュー選択画面に気をとられることなく、手を拘束されずに試着イメージに没頭できることなどが考えられる。また、与えられた選択肢から選ぶだけになると、システムに選ばれているような気分になり、ユーザが主導権を持って自然にシステムと対話を行うのが困難になる。

次に本システムを、大スクリーン上の動画像合成ではなく、通常のコンピュータディスプレイに静止画で表示す

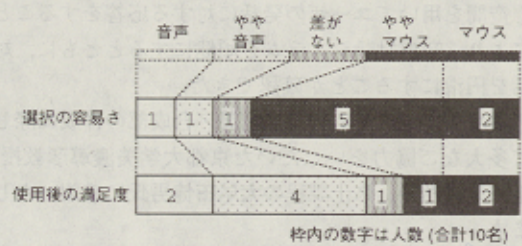


図5 音声入力 vs. マウスによるメニュー選択

Fig. 5 Speech input vs. menu selection by mouse.

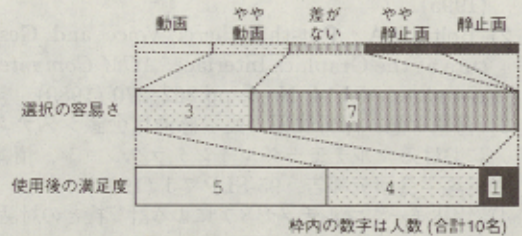


図6 大スクリーン上の動画像合成 vs. 計算機画面上の静止画像合成

Fig. 6 Video image on large screen vs. still pictures on computer display.

るシステムと比較した。同じ10人の被験者に同様に比較評価を行ってもらった。その集計結果を図6に示す。選択の容易さについてはあまり差がないが、満足度に関しては大多数の被験者が動画像合成の方を支持している。計算機ディスプレイ上の静止画として出力する場合、ユーザの試着イメージは写真のようになり、細部が分かりにくいだけでなく、(仮想鏡の前で動作したりできないので)次の試着イメージが合成されるまで待つだけになり、発話のたびにシステムとのインタラクションが中断されるためである。

本タスクは、作図などと異なり、ゴールが明確に規定されないため、対話に要する時間等で有効性を評価できない。むしろ、対話はずみ、ユーザの満足度が高くなるほど望ましいと考えている。以上の結果から、本システムがユーザに高い満足度を与えることが確認された。

5. おわりに

本論文では、仮想空間技術と音声対話技術を統合した新しいヒューマンインタフェースを探るために、仮想空間と音声言語の特徴について考察した。これをもとに、設計・作成した仮想試着システムについて説明した。仮想空間との対話で音声言語を用いることにより、属性やニュアンスによる指示が可能になり、これらが配色の指定などにおいて効果的であった。また仮

想空間を用いてユーザの発話に対する応答をすることにより、音声対話システムを頑健にするとともに、対話を円滑にすることが確認できた。

謝辞 本システムの画像認識・合成部の作成に際して多大なご協力をいただいた京都大学美濃導彦教授、亀田能成助手、修士課程の太尾田健男氏に感謝いたします。

参考文献

- 1) 廣瀬通考：バーチャルリアリティー，オーム社 (1995).
- 2) Bolt, R.A.: Put-that-there: Voice and Gesture at the Graphics Interface, *ACM Computer Graphics*, Vol.14, No.3, pp.262-270 (1980).
- 3) 宮里 勉，岸野文郎：臨場感通信会議システムにおけるマルチモーダルインタラクション，情報処理学会研究報告，95-SLP-7-1 (1995).
- 4) 竹林洋一：マルチメディアによる計算機との対話，電子情報通信学会技術研究報告，SP92-37 (1992).
- 5) 竹林洋一：音声自由対話システム TOSBURG II—ユーザ中心のマルチモーダルインタフェースの実現に向けて，電子情報通信学会論文誌，Vol.J77-DII, No.8, pp.1417-1428 (1994).
- 6) 安藤ハル，北原義典，畑岡信夫：インテリアデザイン支援システムを対象としたマルチモーダルインタフェースの評価，電子情報通信学会論文誌，Vol.J77-DII, No.8, pp.1465-1474 (1994).
- 7) 西本卓也，志田修利，小林哲則，白井克彦：マルチモーダル入力環境下における音声の協調的利用—音声作図システム S-tgif の設計と評価，電子情報通信学会論文誌，Vol.J79-DII, No.12, pp.2176-2183 (1996).
- 8) 神尾広幸，松浦 博，正井康之，新田恒雄：マルチモーダル対話システム MultiksDial，電子情報通信学会論文誌，Vol.J77-DII, No.8, pp.1429-1437 (1994).
- 9) 吉岡 理，荒井和博，管村 昇，嵯峨山茂樹：音声認識機能を含むマルチモーダルインタフェースをもつ住所入力システム開発と評価，電子情報通信学会論文誌，Vol.J80-DII, No.5, pp.1007-1015 (1997).
- 10) Kawahara, T., Lee, C.H. and Juang, B.H.: Key-Phrase Detection and Verification for Flexible Speech Understanding, *Proc. Int'l Conf. on Spoken Language Processing*, pp.681-684 (1996).
- 11) 伊藤慶明，古川 清，中沢正幸，木山次郎，張建新，岡 隆一：複数ユーザによる音声とジェスチャのマルチモーダルインタフェースシステム，情報処理学会研究報告，96-SLP-10-2 (1996).
- 12) 河原達也，松本真治，堂下修司：単語対制御をヒューリスティックとする A*探索に基づく会話音声認識，電子情報通信学会論文誌，Vol.J77-DII, No.1, pp.1-8 (1994).
- 13) 坂口嘉之，美濃導彦，池田克夫：仮想服飾環境 PARTY—人体と型紙のための幾何学的制約充足型格子形成法，電子情報通信学会論文誌，Vol.J77-DII, No.11, pp.2210-2219 (1994).
- 14) 河原達也，北岡教英，堂下修司：A*探索に基づいたフレーズポッティングによる頑健な音声理解，電子情報通信学会論文誌，Vol.J79-DII, No.7, pp.1187-1194 (1996).
- 15) 小林重順，道江義頼：応用色彩心理，誠心書房 (1975).
- 16) 小林重順：配色イメージワーク，講談社 (1995).

(平成 9 年 6 月 27 日受付)

(平成 10 年 1 月 16 日採録)



河原 達也 (正会員)

1987 年京都大学工学部情報工学科卒業。1989 年同大学院修士課程修了。1990 年同博士後期課程退学。同年京都大学工学部助手。1995 年同助教。1995 年から 1 年間米国ベル研究所客員研究員。音声認識・理解の研究に従事。京大博士 (工学)。電子情報通信学会，日本音響学会，人工知能学会，IEEE 各会員。



田中 克明

1997 年京都大学工学部情報工学科卒業。現在，同大学院修士課程在籍。音声対話の研究に従事。人工知能学会会員。



堂下 修司 (正会員)

1958 年京都大学工学部電子工学科卒業。1960 年同大学院修士課程修了。1963 年同博士課程単位取得退学。同年京都大学工学部助手。1965 年同助教。1968 年東京工業大学理工学部助教。1973 年京都大学工学部教授。1996 年京都大学大型計算機センター長 (併任)。その間，音声の分析と認識，オートマトンの学習的構成，自然言語処理，人工知能等，知的情報処理の研究・教育に従事。京大工博。1959 年通信学会稲田賞受賞。1988 年人工知能学会論文賞受賞。1990 年情報処理学会創立 30 周年記念論文賞受賞。人工知能学会前会長。電子情報通信学会，日本音響学会等各会員。