T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

論題(和文)	音声による３次元直接操作インタフェース
Title(English)	
著者(和文)	川崎智久, 大西 翼, 篠崎 隆宏, 古井貞熙
Authors(English)	Tomohisa Kawasaki, Oonishi Tasuku, Takahiro Shinozaki, SADAOKI FURUI
出典(和文)	インタラクション2009, , , p. 43-44
Citation(English)	, , , p. 43-44
発行日 / Pub. date	2009, 3
権利情報 / Copyright	ここに掲載した著作物の利用に関する注意: 本著作物の著作権は（社）情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。 The copyright of this material is retained by the Information Processing Society of Japan (IPSJ). This material is published on this web site with the agreement of the author (s) and the IPSJ. Please be complied with Copyright Law of Japan and the Code of Ethics of the IPSJ if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof.

Powered by T2R2 (Science Tokyo Research Repository)

音声による３次元直接操作インタフェース

川崎智久[†] 大西翼[†]
篠崎隆宏[†] 古井貞熙[†]

Voice-based direct manipulation for 3D interface

Tomohisa Kawasaki,[†] Tasuku Oonishi,[†] Takahiro Shinozaki[†]
and Sadaoki Furui[†]

1. はじめに

近年，計算機性能やコンピュータグラフィックス (CG) 技術の向上により，Google Earth や Second Life などの仮想 3D 空間を用いたアプリケーションが普及しつつある．現在，これらのアプリケーションの操作にはマウスやキーボードを利用するのが一般的である．しかし，マウスは基本的に 2D 平面上での移動量を取り込むデバイスであり，従来の 2D 的なグラフィカルユーザインタフェース (GUI) 操作に対しては非常に優れているが，3D 空間内のオブジェクトを操作しようとすると，必ずしも直感的なインタフェースとはなっていない．また，マウスとキーボードの間を頻繁に行き来する必要が生じる場合も多く，操作が煩雑になる傾向がある．

特に，仮想 3D 空間内のアバターや仮想的な視点カメラを移動させるような高い現実感を伴う場合，マウスやキーボードに代えて人間が普段使用している音声を用いて操作できれば，直感的で自然なインタフェースが実現されると考えられる．また音声を用いた場合，オブジェクトの連続的な移動の指示と特定の操作を表すコマンドを自由に混在させられる利点もある．

そこで本研究では，音声を利用して 3D 空間内のオブジェクトの直接操作を行う音声入力インタフェース「Voice-based Direct Manipulation (VDM) インタフェース」を提案する．音声を用いた提案インタフェースは自然かつ直感的であり，キーボードやマウス操作に不慣れな子供や年配者であっても簡単に 3D アプリケーションを楽しむことができると期待される．

† 東京工業大学大学院 情報理工学研究科
Graduate School of Information Science and Engineering, Tokyo Institute of Technology

2. 提案インタフェースの概要

提案インタフェースでは，ユーザは音声で 3D 空間内のオブジェクトを連続的に移動したり，特定の操作を指示することができる．以下で，各機能について説明する．実装の詳細については，3 節で説明する．

2.1 音声による移動操作

これまでにオブジェクトの移動のような直接操作を音声で行う手法としては，単語発声に基づく五十嵐[1] や Olwal 等[2] によるシステムや，母音発声に基づいた Vocal Joystick[3] が提案されている．しかし，これら従来の単語発声に基づく方法ではオブジェクトの移動が単語の発声が完全に終わるまで始まらないという点でレスポンス性能に問題があり，また母音発声に基づいた手法では母音と方向の対応を覚えなければならないうえ，母音の認識精度が低く思い通りに操作することが難しいという問題が存在した．

このような背景のもと，我々はこれまでに単語発声とその末尾母音の引き延ばしにより方向と移動量を表現し，また高速なデコーダ技術と組み合わせることで単語の発声終了を待たずにオブジェクトの移動が開始される手法を提案したが[4]，これを 3D 空間内のオブジェクト操作に応用・発展させたものが本研究で提案するインタフェースである．提案インタフェースではオブジェクト移動の際，移動させたい方向自体を発声するので動作が直感的に分かりやすく，かつ一回の発声で移動処理を行うことができるので，高い即時性を実現している．例えば「前ーーー」と発声している間，

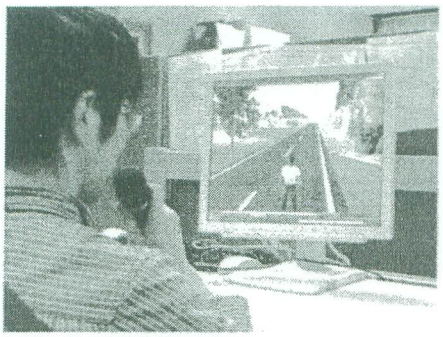

図 1　提案する VDM システム
Fig. 1　proposed VDM system

オブジェクトが前に移動し続け，発声が終了した時点でオブジェクトの移動が終了する．方向は単語を単位とした認識によって決定されるので，認識精度が高く安定した動作が可能となっている．

また本研究では新たに，「もうちょっと」や「少し」などの，程度を示す単語を発声することで細かな移動制御を行う機能を実装した．例えば，「少し右」のように，程度を示す単語の後に方向を示す単語を発声した場合，その方向に対して細かな移動制御を行い，「もうちょっと」のように，程度を示す単語のみを発声した場合，最後に発声された方向を示す単語に基づき移動制御を行う．この機能により，大まかな操作と細かな制御を両立させることが可能となった．

2.2　音声によるコマンド処理

提案インタフェースでは，「つかむ」「離す」などの処理名を発声することによって，その単語に対応したオブジェクト操作を行うことができる．マウスやキーボードでこれらの処理を行う場合，コマンドやメニューを用いて間接的に行うのが一般的であるが，我々が提案するインタフェースでは，同様の処理を音声で直接的に行うことができる．

3. 実　　装

提案インタフェースを Second Life に実装したシステムを図 1 に示す．Second Life の仮想 3D 空間内では，アバターを操作し，歩いたり，飛んだり，ものを売買したりすることができる．Second Life ではキーボードやマウスを用いた操作の複雑性が指摘されているが，本インタフェースを用いることで直感的な操作が可能となる．

ユーザが発声した音声を機械に理解させるための音声認識デコーダには，我々が開発を行っている T^3 デコーダ[5] を用いた．T^3 デコーダでは，早期に単語を出力するための手法として，仮説の共通単語履歴の検出に基づく手法[6] を採用している．この手法では発声途中に単語列を出力することが可能となるため，発声単語の語尾を引き延ばし続けても，発声の終了を待つことなく認識結果を得ることができる．本システムでは「上」「下」「左」「右」「前」「後ろ」の 6 方向を示す単語と，「もうちょっと」「もう少し」「あとちょっと」「あと少し」「ちょっと」「少し」「もっと」の，程度を示す 7 つの単語，及び「つかむ」「離す」「座る」「立つ」の 4 つの処理名を認識することができる．

4. システム評価

提案インタフェースを用いて実装した上記システムを，「信学会・子供科学教室」の場で小学生数十人に対し展示しデモを行い，評価してもらった．デモを見た小学生の多くが，3D 空間内のオブジェクトを音声で操作するという今まで行ったことがない操作に目新しさを感じ，好意的な評価をした．Second Life 上に実装したことで，小学生達はゲーム感覚で本システムを楽しめたらしく，展示終了後に「とても面白い」「是非使ってみたい」などの感想を多く得た．

5. おわりに

本研究では，音声を利用して 3D 空間内のオブジェクトを直接操作できる入力インタフェースを提案し構築した．今後の課題としては，音声認識デコーダに耐雑音技術を実装し，また未知語処理を行うことで周辺雑音に対する頑健性を向上させること，オブジェクト処理の機能を更に増やすことでより自由度の高いシステムとすることなどが挙げられる．さらに，提案インタフェースを実世界のロボットの直接操作に応用し，人間とロボットとのより自然なインタラクションの実現を目指す．

参 考 文 献

1) 五十嵐健夫ほか：言語情報を用いない音声による直接操作インタフェース，*Proc. WISS* (2001).
2) Olwal, A., et al.: Interaction Techniques Using Prosodic Features of Speech and Audio Localization, *Proc. IUI*, pp.284–286 (2005).
3) Jeff, A. Bilemes., et al.: The Vocal Joystick, *Proc. ICASSP*, Vol.1, pp.625–628 (2006).
4) 川崎智久ほか：音声入力によるマウスの直接操作の検討，秋季音講論，pp.55–56 (2008).
5) Paul, R. Dixon., et al.: The TITech large vocabulary WFST speech recognition system, *Proc. ASRU*, pp.443–448 (2007).
6) Brown, P., et al.: Partial traceback and dynamic programming, *Proc. ICASSP*, pp.1629-1632 (1982).