論文 / 著書情報
Article / Book Information

論題(和文)	音声入力によるマウスの直接操作の検討
Title(English)	Voice-based direct manipulation interface for mouse cursor control
著者(和文)	川崎 智久, 大西 翼, 岩野 公司, 篠崎 隆宏, 古井貞熙
Authors(English)	Tomohisa Kawasaki, Oonishi Tasuku, Koji Iwano, Takahiro Shinozaki, SADAOKI FURUI
出典(和文)	日本音響学会2008年秋季講演論文集, , No. 1-1-23, p. 55-56
Citation(English)	, , No. 1-1-23, p. 55-56
発行日 / Pub. date	2008, 9

音声入力によるマウスの直接操作の検討*

◎川崎 智久，大西 翼，岩野 公司，篠崎 隆宏，古井 貞熙 (東工大)

1 はじめに

現在，パソコンやワープロなどの情報処理機器へのアクセスには，文字入力はキーボード，選択操作はマウスを手や指で操作して利用するのが一般的である．しかし，手に障がいがある場合や両手が塞がれている場合など，これらのデバイスを手や指を使って自由に操作することが困難な場合も存在する．このような場合，これらのデバイスを手を使わず操作できるような補助機能や代替機能／機器が必要となる．そのため近年，マウス操作を行う音声入力インタフェースの研究がなされている [1, 2]．しかし従来方式は，インタフェースとしての直感性や即時性が十分とは言えず，改善の余地がある．そこで，本研究では，これらの問題をふまえ，音声を利用してマウスの直接操作を直感的かつ即時的に行う音声入力インタフェースを提案する．

2 従来研究の概要と問題点

音声によるマウス操作を実現した例として，五十嵐 [1] のシステムと Vocal Joystick[2] が挙げられる．五十嵐のシステムでは，例えば,「上へ移動，あーーー」と発声すると「あーーー」と発声している間カーソルが上へ移動し続ける．しかし，このシステムでは,「上へ移動」という操作決定のための発声と，実際の操作のための「あーーー」という発声の二段階の入力を必要とし，即時性の面で問題がある．一方，母音認識を利用した Vocal Joystick では,「æーーー」と発声している間，/æ/ に対応する上方向にカーソルが移動を続ける．しかし，このシステムを操作するためには，母音と方向の対応を覚えなければならないため，直感性の面で問題がある．また，即時性を向上させるため，フレーム毎に母音の識別を行っているので，認識精度が低く，動作が不安定であるという問題もある．

これらの問題をふまえ本研究では,「上ーーー」のように，方向を示す単語の語尾を引き延ばし発声し続けることで目的の方向にカーソルを移動させ続ける，直接操作性の高い音声入力インタフェースを提案する．提案手法では，カーソル移動の際，移動させたい方向自体を発声するので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，即時性も高い．また，方向は単語を単位とした認識によって同定されるので，フレーム単位で識別する方法に比べ認識精度も高く，安定した動作を示す．

3 システムの基本設計

本研究で提案するインタフェースを実現するシステムの概要図を Fig.1 に示す．本システムは大きく分けて以下の 3 つの部分から成り立っている．

- 発話区間検出部
- 音声認識部
- マウス動作制御部

ユーザが発声した音声はまず発話区間検出部と音声認識部に送られる．発話区間検出部では，受け取った音声から発声の有無を検出し，発話区間情報を出力する．音声認識部では，受け取った音声を認識し，

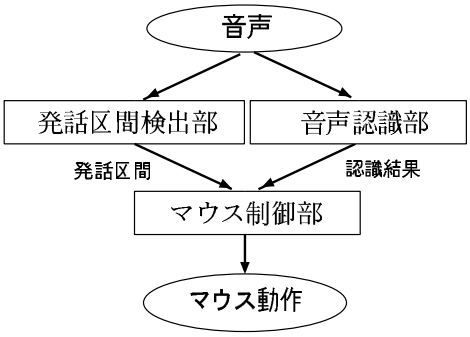

Fig. 1 システムの概念図

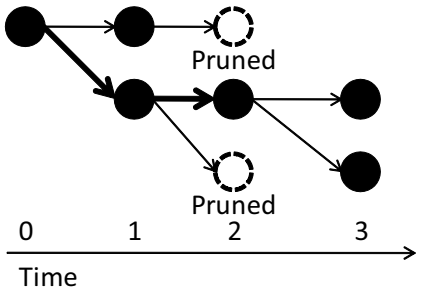

Fig. 2 共通履歴の検出による早期仮説確定

その結果をマウス制御部へ送る．マウス制御部は，認識結果に対応するマウス動作を行う．以下で，各部分についての説明を行う．

3.1 音声認識部

音声認識部には，東京工業大学で開発が行われている WFST(Weighted Finite State Transducer) に基づく音声認識デコーダ [3] (「T^3 デコーダ」) を用いた．T^3 デコーダの探索は，フレーム同期型の 1 パス探索であり，早期に確定した仮説単語列を出力することが可能である．

Fig.2 に，共通する単語履歴の検出に基づく早期仮説確定の例を示す．図は，時刻 $t = 0 \sim 3$ における探索を示しており，黒丸は探索ネットワーク中の各時刻におけるアクティブな状態を表している．つまり，この図では $t = 3$ の時点で 2 つの仮説が生き残っている．$t = 0 \sim 2$ の太線のパスは，$t = 3$ で生き残っている全ての仮説に共通する履歴となっており，この部分は $t = 3$ 以降の探索でも変化しない．このような共通履歴部分が確定した段階で，該当する単語系列の出力を行う．この手法は，最終的な認識結果に影響を与えないため，認識率の低下を起こさずに，早期に単語列を確定することが可能である．

この認識部では,「上」「下」「左」「右」「左上」「左下」「右上」「右下」の 8 方向を示す単語と「クリック」「ダブルクリック」「開く」「閉じる」「コピー」「はりつけ」「つかむ」「離す」の 8 つの処理名を認識することができる．

3.2 発声区間検出部

上記で説明した手法により，発声単語の語尾を引き延ばし続けても，発声が終了する前に，認識結果を得

*Voice-based direct manipulation interface for mouse cursor control by Tomohisa Kawasaki, Tasuku Oonishi, Koji Iwano, Takahiro Shinozaki and Sadaoki Furui (Tokyo Institute of Technology)

ることができる．一方，語尾の引き延ばしが終了した時点は検出することはできないため，別途対応が必要となる．そこで本システムでは，発声の有無を検出する発話区間検出部により，単語の語尾の引き延ばし判定を行う．

実際には，発話区間検出部はデコーダのフロントエンドとして実装されており，零交差法を利用して，発話区間の検出を行っている．

3.3 マウス制御部

音声認識部から受け取った認識結果が方向を示す単語である場合，発話区間検出部から得られる発話区間情報と組み合わせ，方向を示す単語の語尾を引き延ばし発声し続けている間，その方向にカーソルを移動させ続ける．一方，認識結果が処理名である場合，マウス制御部は処理名に対応する操作を行う．これらのマウスの制御は Windows API を利用して行った．

4 評価実験

4.1 実験条件

提案したシステムを PC に実装し，実験に用いた．OS は Microsoft Windows XP Professional Version 2002 Service Pack 2，CPU は Intel Pentium M(1.3GHz)，内蔵メモリは 1GB となっている．

実験は研究室内で行い，システムを操作する際には，音声入力用ハンドマイクを使用してもらった．被験者は男子大学 (院) 生 10 名である．

4.2 実験手順

被験者に対しては予め，システムの操作方法，及び発話可能単語 (8 方向+8 処理) を説明してから実験を行った．システムの操作方法が分からなくなった被験者には，その場で助言を与えた．

実験は，フォルダの開閉やファイルの移動など実際に手が使用できない状況下でマウスを操作する場面を想定して行った．具体的には，画面中央にあるフォルダから画面端にあるフォルダ，ファイルへカーソルを移動させ，コピー操作などを行い，中央のフォルダにカーソルを戻し，処理を終了させる．被験者はこの一連の流れを 8 方向に対し行う．

実験終了後，被験者には以下の 8 項目について 5 段階 (1. とても悪い ~5. とても良い) で主観的に評価してもらった．

1. 直感性：操作は分かりやすいか
2. 移動開始時の即時性：方向を示す単語を発声してから実際にカーソル移動が開始するまでの遅延は小さかったか
3. 操作開始時の即時性：処理名を発声してから実際に対応する操作が行われるまでの遅延は小さかったか
4. 移動終了時の即時性：方向を示す単語の発声を終了してから実際にカーソル移動が終了するまでの遅延は小さかったか
5. 認識性能：発声したとおり正しくマウス動作は行われたか
6. ストレス：このインタフェースを使ってみて感じたストレスは少なかったか
7. 利便性：手の使えない状況でこのインタフェースは便利であるか
8. 総合評価

また，上記項目 2~4 に対応する各遅延時間 (T2, T3，T4) を各被験者ごとに測定し，それらの平均値を客観データとして用いた．

Table 1　主観評価結果

1. 直感性	4.8	5. 認識性能	4.6
2. 移動開始時	4.7	6. ストレス	3.4
3. 操作開始時	4.8	7. 利便性	4.0
4. 移動終了時	4.6	8. 総合評価	4.4

Table 2　発声とマウス動作のタイミングの遅延

T2(秒)	T3(秒)	T4(秒)
1.51	1.76	0.60

4.3 実験結果

被験者による主観評価結果を Table 1 に，客観評価結果を Table 2 に示す．

Table 1 を見ると,「直感性」について高い評価が得られている．これは，今回構築したインタフェースが，直感的で分かりやすいとユーザに認められた結果であると考えられる．

また「移動開始時」「操作開始時」「移動終了時」の各即時性に対しても，高い評価が得られている．これは，ユーザが発声と実際のマウス動作のタイミングの遅延を小さいと感じたことを意味している．実際の遅延時間は，Table 2 を見ると，発声開始からマウスが動作し始めるまでの遅延時間は 1.5~2 秒程度，方向を示す単語の語尾の引き延ばし発声の終了判定からカーソル移動が終了するまでの遅延時間は 0.6 秒であり，この程度の遅延が，実用上問題ない程度に小さいことが示された．

「認識性能」の評価も高い値を示した．本システムでは雑音処理，未知語処理を行っていないため，雑音が誤って認識されたことで生じる挿入誤りがあり，今回の実験におけるキーワード認識精度は 87.7 %であった．しかし，ユーザが意図して発声したキーワードは 100 %正しく認識されているため，この「認識性能」の評価が高かったものと考えられる．

「ストレス」の項目の評価は全体に対して低い値を示した．これは，音声を連続的に出し続けなければならないことが，ユーザにとって負担となってしまったことが大きな原因だと考える．

5 おわりに

本研究では，音声を利用してマウスを直接操作できる入力インタフェースを構築した．被験者実験を行い，発声とマウス動作のタイミングの遅延が実用上問題ない程度に小さいことが示された．また，主観評価においては,「直感性」などの項目において高い評価を得た．以上より，本研究で構築したインタフェースは，直感的かつ即時的な直接操作性の高いインタフェースであることが示された．

今後の課題としては，雑音処理，未知語処理を新たにシステムに含め，キーワード認識精度を向上させる必要がある．また，単語の語尾の引き延ばしにかかるユーザへの負担を軽減する努力も必要である．

参考文献

[1] 五十嵐 健夫，John F. Hughe, “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*, 2001.

[2] Jeff A. Bilemes, et al., “The Vocal Joystick,” *Proc. ICASSP 2006*, vol.1, pp.625–628, 2006.

[3] Paul R. Dixon, et al., “The TITech large vocabulary WFST speech recognition system,” *Proc. ASRU 2007*, pp.443–448, 2007.