

## マウス操作を模擬した音声カーソル制御\*

田上集<sup>†</sup> 光本大記<sup>‡</sup> 嵯峨山茂樹<sup>‡</sup>

<sup>†</sup>明治大学 総合数理学部 / <sup>‡</sup>明治大学 先端数理科学研究科

### 1. はじめに

本稿では、音声特徴量によるカーソル操作の方式3種を比較検討する。

多くのユーザインターフェースにおいて、マウスカーソル(ポインタ)の移動とクリック操作が多用されるが、料理中や車の運転中で両手が塞がっている場合や手に障害がある場合などでは操作が困難である。手に代わる音声によるカーソル操作の研究では、五十嵐らは、発声の有無によるオン・オフ操作、音程によるパラメータとタンギングによる離散的パラメータの増減操作を組み合わせる手法を提案した[1]。Bilmesらは、あらかじめ対応する方向を設定された母音が入力音声から認識されている間、その方向へのカーソル移動が行われる Vocal Joystick を提案した[2]。川崎らは「うえーー」「みぎーー」など方向を示す単語の語尾を引き延ばす発声で目的の方向にカーソルを移動させ続ける、音声入力インターフェース手法を提案した[3]。

### 2. 音声特徴量によるカーソル操作

快適なカーソル操作のためには、ユーザの発声が即時にカーソル移動に反映されること、カーソル移動方向が制約を受けないこと、ユーザ登録が不要なことなどが望ましい。しかるに音声認識は、発声開始から認識結果を得るまで遅延を生じること、移動方向が単語に対応した離散的方向数に制約されること、場合によっては話者登録が必要になるなど、不利な点がある。

それゆえに我々は、言語音声の認識を用いずに、音声の「音」としての特徴として、母音の第1、第2フォルマント周波数  $F_1, F_2$  が形成する日本語母音の五角形を画面にマッピングすることを検討した。しかし、フォルマントは頑健な抽出や時間軌跡としての連続性に難点があるため、マウスカーソル操作には適さない。フォルマントと似た挙動と連続性を持つ CSM 周波数値[6]で代用し、射影変換して画面上の点  $p = (x, y)$  にマッピングした[4]。

さらに、ポインタの上下位置は声の基本周波数  $F_0$  の高低として直感的に理解し易いと考え、 $y$ 座標を  $F_0$ 、 $x$ 座標を  $F_1$  と関連付けたカーソル操作を検討した[5]。

### 3. カーソルの位置制御方式とクリック制御

#### 3.1 手動カーソル操作の方式の分類

手により行われるポインティング操作の方式には少なくとも以下の3種類がある。

- (A) タッチパネル、タッチペン、タブレット、画面上の点に直接アクセスする方式
- (B) ジョイスティック、トラックポイントなど点移動の

速度ベクトルを与える方式

- (C) マウスポインタ、トラックボール、タッチパッドなど移動ベクトルを継ぎ足す方式

手による操作を声に置き換えるときにも、以上の異なる方式の操作が考えられる。我々は、音声特徴量の上記の

(A)と(B)に対応する操作方式を検討し、即時性や方向の柔軟性を重点にして取り組んだ[5]が、可動範囲のキャリブレーションのために、事前にユーザの音声特徴量を採取し、記憶させなければならなかった。

本稿では、その問題の解決法としてマウス操作を模擬した手法を追加検討し、既方式2種と比較検討する。以下では、 $v = (F_1, F_0)$  を音声ベクトルと呼ぶことにする。

#### 3.2 2次元平面上で画面の目標座標に直接移動する方式 (Type A)

タブレット、スマートフォン画面、タッチペンなどを模した手法。これらはタッチした位置がそのままポインタ座標になる。本研究ではキャリブレーションにより基本周波数の上限値と下限値を測りその数値を  $x$  又は  $y$  座標の上限と下限に設定し、音声ベクトルを画面の位置ベクトル  $p$  に変換する方式である。即ち、

$$p = av + b$$

で表せる。ここで、 $a$  はスカラー係数、 $b$  は定数ベクトルである。

#### 3.3 $(x, y)$ 値をカーソル移動速度ベクトルとする方式 (Type B)

トラックポインタ、ジョイスティックなどを模した手法。本研究では、Type A で利用したキャリブレーションにより特徴量の上限値と下限値の中間値を原点とし、即ち、

$$dp/dt = av + b$$

で表せる。

#### 3.4 マウス操作を模した移動量継ぎ足し方式 (Type C)

マウス、トラックボール、タッチパッドなどを模した手法。この方式では、発声直後のカーソル位置を始点とし、その後の音声特徴ベクトルの変化量(の定数倍)をカーソルの相対値とするものである。即ち、

$$p = p_0 + \sum_{i=1}^n \Delta v_i$$

で表せる。ここで、 $p_0$  は始点位置、 $\Delta v_i$  は  $i$  回目の発声の音声ベクトルの変化量である。発声を繰り返せば、カーソル移動量は継ぎ足せるので、複数回の繰り返し発声により、一度の発声では到達できない距離の移動も容易である。これにより、話者の音声特徴量のキャリブレーションを不要とすることができる。

#### 3.5 無声摩擦音/s/パターンによるクリックとドラック&ドロップ操作

声の特徴量は音韻性、音高のほかパワーや継続時間が提案されたが、パワーはマイクとの距離に影響するため不適切である。そこで本研究では無声摩擦音/s/を利用した。短い/s/の発声時にクリック、長い/s/発声時ドラックし、連続的に発声することでドラックしたままカーソル操作を行える。発声が終了時、カーソルの位置にある座

\*"Voice-controlled cursor inspired by manual mouse operation" by Tsudoi Tanoue, Daiki Mitsumoto, Shigeki Sagayama, Meiji University

標にドロップする手法を検討した。

## 4. 本方式の比較実験評価

### 4.1 実験方法と方式比較

音声によるカーソル操作システムの有用性を評価するため、Fitts の法則を用いて実験を行った。比較対象は Type A, B, C の 3 方式とした。実験フェイズでは、試行ごとに画面上にランダムな位置と大きさで表示される円を、3 種類の提案手法を用いて被験者にクリックしてもらい、操作時間の計測を、3 手法各 50 回の試行について、15 名の被験者に対して行った。カーソルの現在位置から円の中心までの距離  $D$ 、円の直径  $W$ 、カーソル移動に要した時間  $T$  を、Fitts の法則を用いて比較した。Fitts の法則は、インターフェースの指標になり、目標までの移動に要する時間  $T$  は、オブジェクトのサイズ  $W$  と移動距離  $D$  であらわされる困難度指標に依存するという法則であり、

$$T = a + b \log_2 (D/W + 1)$$

で表現され、カーソルを動かすとき、オブジェクトまでまうすお移動する距離が長くなるほど労力がかかり、オブジェクトが小さければ、到達することが困難となることを意味し、定数  $a, b$  の値によりインタフェースとしての使いやすさを評価できる。

また、計測を開始する前に 10 分程度の練習時間を設けた。

### 4.2 実験結果

独立変数を操作方式、従属変数を実験フェイズから得たデータから Fitts の法則を用いて評価を行なった回帰直線の傾きとして分散分析を行った結果、操作速度に関しては Type A と Type B の間には有意差が見られなかったが、ともに Type C より有意に速いことが危険率 5% で示された。

### 4.3 考察

被験者からは、Type A と Type B は「思った方向に難なく動かせる」や「ストレスがない」等の感想が得られ、事実比較的短時間で使いこなしている被験者が多かった。一方、Type C 方式は「操作方法が覚えにくい」といった意見が多く、使いこなすのに時間がかかった。これは他 2 方式と比べ移動方法が少し複雑であることが被験者に抵抗があったと考えられ、十分に慣れていないと調整や操作しにくい。しかし、「操作速度が調整できればさらに使いやすくなる」や「慣れれば最も使いやすいのでは」という意見もあった。また、今回の実験では速度を重視した評価傾向であったため、細かな操作に適していると考えられる Type C については再度検討が望ましい。

## 5. 結論

カーソル操作を声の特徴量を用いて 2 次元平面上で画面の目標座標に直接移動する方式 (Type A)、 $(x, y)$  値をカーソル移動速度ベクトルとする方式 (Type B)、マウス操作を模した移動量継ぎ足し方式 (Type C) の 3 通りの操作方法を比較検討した。結果、Type A、Type B 方式は、Type C 方式より有意に早いことが示された。

今後の課題としては、カーソル速度制御、ホイール操作や右クリックの検討を行いたい。なお、Type C については今回行った実験では優位性を示すことができなかったが、事前のキャリブレーション不要な点や、速度を調整すれば微細な動きにより高い分解能の実現可能性があるという点から、その面を重視した実験評価を行っ

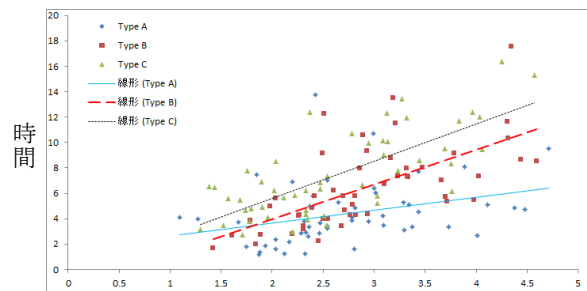


Fig.1 被験者一人の実験結果

ていきたい。

### 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ーフェース,” Proc. WISS 2001, pp. 81-82, 2001.
- [2] Jeff A. Bilmes, et al., “Vocal Joystick,” Proc. ICASSP2006, Vol. 1, pp. 625-628, 2006.
- [3] 川崎智久, 大西翼, 岩野公司, 篠崎隆宏, 古井貞熙, “音声入力によるマウス直接操作検討,” 日本音響学会 2008 年秋季講演論文集, 1-1-23, pp. 55-56, 2008.
- [4] 田澤健斗, 篠原一輝, 嵯峨山茂樹, “CSM 分析を用いた音声マウスカーソル操作,” 2016 年度情報処理学会全国大会講演論文集, 5Z-02, Mar. 2016.
- [5] 田上集, 光本大記, 小口純矢, 松原聖人, 嵯峨山茂樹, “CSM 周波数と基本周波数を用いたマウスカーソル操作,” 2017 年度情報処理学会全国大会講演論文集, 4ZA-02, Mar. 2017.
- [6] 嵯峨山茂樹, 板倉文忠, “複合正弦波モデルによる音声分析,” 電子通信学会論文誌, Vol. A64 (2), pp. 105-112, 1981.