5Z-02

CSM 分析を用いた音声によるマウスカーソル操作*

田澤 健斗†　篠原 一輝†　嵯峨山 茂樹†

(明治大学総合数理学部)

1. はじめに

調理中にレシピのページを変えたい時や、車の運転中など、手を使わずにマウス操作を行いたい場面がある。また、手に障害があり両手が不自由なためにマウス操作が困難なこともある。本研究では、そのような環境でマウス操作を行うために、音声によるマウス操作の実現を目標とする。

音声認識を用いる従来の手法とは異なり、二次元平面上の任意の点に直接移動する方法を検討した。音声から得られる二次元量としては音声フォルマント周波数が考えられるが、時間連続性に難がある。そこで、それに性質が近く時間連続性を持つ値を複合正弦波モデル(Composite Sinusoidal Modeling, CSM)分析を用いることを検討した。

2. 音声によるマウスカーソル制御の原理と理論

2.1 従来研究の概要と問題点

音声を用いるマウス操作の研究としては、発声の有無によるオン・オフ操作、音程によるパラメータの増減操作、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方法[1]、音声認識を活用して「上に移動、あーー」などと方向を指定して、その後母音を発している間その方向に移動する方法[2]、Vocal Joystick [3]のシステムなどがある。[3]のシステムでは、「æーー」と発する間/æ/に対応する方向にマウスが動く。いずれの方法も方向を指定してマウスを動かすという方法であるが、いずれもその方向にしか動かせず、いずれも目的の場所にカーソルを動かす点で即時性の面で改善の余地があった。

2.2 音声特徴から画面座標へのマッピング

これらの問題を踏まえて、本手法では即時性を向上させるため、音声のある特徴を座標平面に対応させ、目的の場所への移動を瞬時移動できる可能性があるインタフェースを検討した。音声に座標平面に対応させるため音声の特徴を用いて二つの値を得る必要がある。その目的には、まず第一に、音声のフォルマント(特に第 1 フォルマント、第 2 フォルマント)が考えられる。日本語においては、5母音の第 1、第 2 フォルマント周波数を座標として平面上にプロットすると「母音の五角形」が見られることは良く知られている。

しかし、フォルマントは抽出の困難さ(曖昧さ)と、時間軌跡としての連続性に難点があり、マウス操作には適さない。そこで、連続性があり音声フォルマントに似た特徴を持つ CSM 分析周波数によって得られる CSM 周波数値の利用を検討した。

また、[1,2,3]のいずれのシステムでもクリックの実装はなされてない。本研究では音声パワーを用いたクリック操作の実装を検討した。

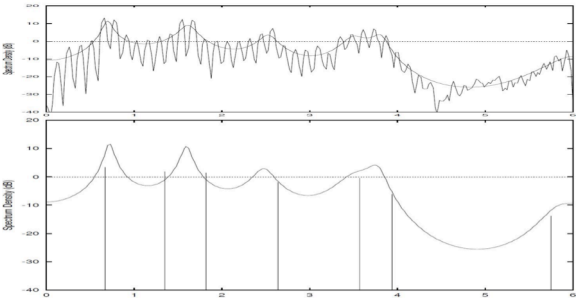

Figure 1. CSM 音声分析の例。上：男声/a/の短時間スペクトル、下：LPC 推定全極型スペクトル曲線と、CSM 周波数と強度（横軸は周波数 0 - 6kHz）

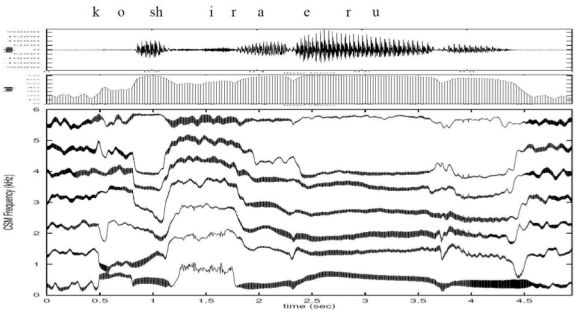

Figure 2. CSM 音声分析の時間軌跡の例。上から、音声波形、パワー、CSM 周波数と強度(線幅)

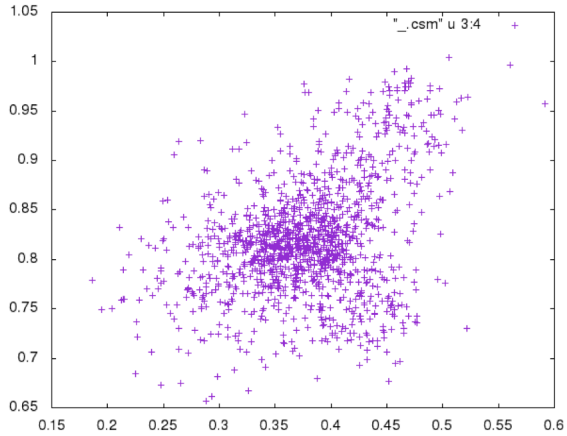

Figure 3. 母音発音の第 1, 第 2 CSM 周波数の分布例

*Voice-controlled mouse cursor through CSM speech analysis
† Kento Tazawa, Kazuki Shinohara, Shigeki Sagayama (Meiji University)

2.3.複合正弦波モデル(CSM)音声分析の性質

CSM 分析[4]は、異なる周波数と振幅の複数の正弦波を重畳したモデル（複数線スペクトルモデル）、その自己相関関数の低次の項が、音声の短時間自己相関関数に一致する場合に、モデル正弦波の周波数（線スペクトル周波数）と強度を代数演算によって求める方法である。携帯電話など音声情報圧縮で広く用いられている LSP 音声分析の別定式化ともなっている。

Fig. 1 は/a/の1フレーム(30mS)の音声信号の CSM 音声分析の結果である。CSM 周波数は、必ずしもフォルマントに一致しないが、似た挙動をすることが見られる。また、Fig. 2 は、CSM 周波数と強度の時間軌跡の例であり、少なくとも1次と2次の CSM 周波数は滑らかに変化していることが分かる。

3. システムの基本設計

3.1 平面射影変換による整形と校正

1,2 次の CSM 周波数の分布を調べると、Fig. 3 のように必ずしも長方形ではない。これは「母音の五角形」としても知られていることである。そこで、1,2 次 CSM 周波数の分布を対象画面に一対一にマッピングするためには、座標(x, y) から別の座標(u, v) へ変換する、平面射影変換が利用できる。変換式は、

$$u = (ax + by + c)/(gx + hy + 1)$$
$$v = (dx + ey + f)/(gx + hy + 1)$$

であり、四隅の (x, y) と (u, v) の対応を与えれば、これら変換係数（変換行列要素）$a \sim h$ が求まる。

また、CSM 周波数値は話者に依存するので、まずは変化範囲の校正を行う必要がある。あらゆる母音を発声し、CSM 座標値 (x, y) を見て、画面の四隅に対応する4点を選ぶことによって行える。このようにして、話者登録段階で校正を行える。

3.2 座標値時系列のスムージング

CSM 周波数は分析窓の周期で実時間で得られるが、分析フレーム周期と音声のピッチ周期は基本的には一致しないため、その軌跡は noisy であり、また発声者は安定した発音を持続できない場合もある。そこで、ここでは座標ベクトル$\boldsymbol{x}(t)$の時系列に一次遅れ系による平滑化:

$$\tilde{\boldsymbol{x}}(t) = a\,\tilde{\boldsymbol{x}}(t-1) + (1-a)\,\boldsymbol{x}(t)$$

を行った。但し、

$\tilde{\boldsymbol{x}}(t)$: 時刻 t におけるカーソル位置ベクトル
$\boldsymbol{x}(t)$: 時刻 t における入力された位置ベクトル
a : 適当な定数（$0 < a < 1$）

3.3 マウスクリック操作

音声によるマウスクリック操作のためには、マウスカーソル移動に使用する音声特徴量とは別の特徴量であるパワーやピッチなどが利用できる。本研究では、音声パワーの利用を検討した。

長い発声時に CSM 分析によるカーソル座標移動を行い、短い発声時にそのパワーパターンによってクリック操作を行うこととした。

3.4 動作確認検証

上述の基本設計に従い、プログラムを実装し、動作確認を行った。手順は、様々な母音音声を録音して CSM 分析をし、1,2 次の CSM 周波数を組み合わせた(x, y) 座標データの散布図(Fig.3)を表示させ、それに基づいて四隅の 4 点を適当に選び、平面射影変換係数を求めるものである。

実験は手が使えない状況でファイルの開閉を行う場面を想定して行った。具体的には Windows PC のウインドウ上に4つのファイルをそれぞれ右上、右下、左上、左下に置き、それらの開閉を行ってもらった。動作確認テストは開発者を含めた学生 3 名で行った。

このインタフェース操作には、慣れるまで時間がかかったが、3 名ともファイルの開閉は問題なく行え、動作が検証できた。

4. 結論

音声が持つ性質を利用し、マウスカーソルを画面の目標座標に直に移動できる方式を、CSM 分析を用いて検討した。また、音声のパワーを利用したマウスクリック操作の実現も検討した。さらに、これらの検討に基づいたプロトタイプシステムを実装し、動作確認を行った。

今後は、音声の特徴量をカーソル移動の速度ベクトルとみなす方式も検討し、それぞれの方式のヒューマンインタフェースとしての優位性などを比較評価していきたい。

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フェース,” Proc.WISS 2001, 2001.
[2] 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の直接操作の検討,”日本音響学会 2008 年秋季講演論文集, No.1-1-23, p.55-56, 2008.
[3] Jeff A. Bilmes, et al., “Vocal Joystick,” Proc.ICASSP 2006, vol.1, pp.625-628, 2006.
[4] 嵯峨山茂樹, 板倉文忠, “複合正弦波モデルによる音声の分析,” 電子通信学会論文誌 A 64(2), p105-112, 1981.