

CSM 周波数と基本周波数を用いた音声によるマウスカーソル操作*

田上集[†] 小口純矢[†] 光本大記[†] 松原聖人[†] 嵯峨山茂樹[‡](明治大学総合数理学部[‡])

1. はじめに

パソコンの操作においては、マウスカーソル移動とクリック操作が多く使われる。しかし、料理中や車の運転中などの両手が塞がった状況や、手に障害があり不自由な状況もあり得る。本稿は、そのように手が使えない状況においてもマウスカーソル操作を行えるように、音声によるマウスカーソル操作の実現を目標とする。

2 音声によるマウスカーソル制御の原理

2.1 従来のアプローチと問題点

従来の音声によるマウスカーソル操作では、様々なインタラクションの方法が研究された。例えば、左右上下のカーソル移動方向の発話を音声認識し、その後も継続して母音を発している間、同じ方向に移動を継続する方法がある[1]。

また、Vocal Joystick [2]のように「ee——」と発声している間、/ee/に対応する方向にマウスカーソルを動かす方法などがある。

いずれの方法も指定された方向にしか移動ができない点、また、移動先目標座標によっては、音声認識を何度か用いなければならない点において改善の余地があった。

2.2 音声スペクトル特徴量によるカーソル制御

このような即時性の問題意識から、本研究では、昨今多く使われる言語音声の認識によるのではなく、音声の「音」としての特徴を利用して、母音発声の音声フォルマント周波数と、声の高さに対応する基本周波数を用いて解決することを試みた。実際には、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。そこで、連続性と音声フォルマントと似た特徴を兼ね備えた CSM 周波数値[4]の利用を検討した。

CSM 周波数の第 1 周波数 F_1 と第 2 周波数 F_2 、そして基本周波数 F_0 を使い、

1. 第 1 周波数 F_1 と第 2 周波数 F_2
2. 第 1 周波数 F_1 と基本周波数 F_0
3. 第 2 周波数 F_2 と基本周波数 F_0

* Voice-activated mouse cursor using CSM frequency and fundamental frequency

[†] Tsudoi Tanoue, Junya Koguchi, Daiki Mitsumoto, Kiyoto Matsubara, Shigeki Sagayama

[‡] Meiji University, Dept. of Interdisciplinary Mathematical Sciences

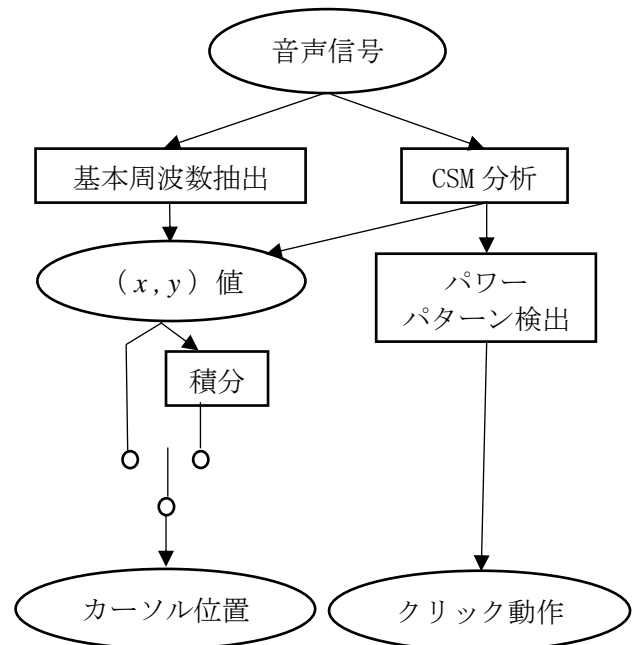


Fig. 1 システムの概念図

のいずれかの組み合わせを (x, y) 値に対応させ、さらに、マウス操作を

- ・二次元平面上で画面の目標座標に直に移動する方式
- ・ (x, y) 値をカーソル移動速度ベクトルとする方式

と組み合わせ、合計 6 通りを検討し比較した。

2.2 複合正弦モデル (CSM)

CSM 分析は、異なる周波数とそれぞれの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）により、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ように、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. 2, 3 に分析例を示す。

3. 本方式の設計と動作実験

3.1 システム設計

本研究では、 x 座標に CSM 周波数の第一フォルマントと第二フォルマントを、 y 座標に基本周

波数を取り入れた。それを実現するためのシステム概要図を Fig.1 に示す。

ユーザーが発声した音声は、まず CSM 分析とフーリエ変換がなされる。CSM 分析によって得られた音声の CSM 周波数を x 座標に、パワーをクリックに利用する。次に、フーリエ変換によって得られたスペクトルをケプストラム法により基本周波数を抽出し、それを y 座標に利用する。抽出された x, y 座標をマウス制御部で先述した 2 通りの方法で定め、マウス操作を実行する。

そのほかに、座標直時系列のスムージング、フレームシフト、入出力レベルを安定させるためにキャリブレーションを導入した。

4.1 動作確認と方式比較

上述の基本設計に従い、プログラムを実装し開発者を含む 3 名で実験を行った。実験は、手が使えない状況での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 のウィンドウ上に 4 つのファイルをそれぞれ右上、右下、左上、左下に配置しそれらの開閉を行った。

実験前は、直観性において不十分さを予想していたが、実際に実行してみると、方向の自在性や移動のコツをつかむことにより比較的使いやすいように感じられ、繰り返すたびに開閉に所要する時間は短くなり、有用可能性を検証できた。3 名ともファイルの開閉は行えた。

予備評価実験として、上述した 6 通りの比較を行ったところ、いずれも実験目標は特に問題なく達成できたが、比較的《第一 CSM 周波数と基本周波数》を用いて《画面の目標座標へ直に移動する方式》が最終的に最も使いやすい、という予備実験結果を得た。その理由としては、 y 座標値はピッチ周波数による声の高さによって決まるので高い位置を高い声として直感的に感じられ、比較的操作しやすい点が考えられる。

また、特徴量を速度ベクトルとして扱う方式は、最初は最も使いやすく、比較的早く実験目標を達成したが、他方のマウス制御方法と比べカーソルの方向に予測がしづらく、実験を何度も行ううちに、僅差ではあるが時間が掛かった。

4. 結論

マウスカーソル操作を CSM 周波数と基本周波数により行う方式を検討した。また、それらを移動速度ベクトルとして扱う方式の検討、キャリブレーションの導入を行い、これらを用いたプロトタイプシステムを実装し動作確認を行った。

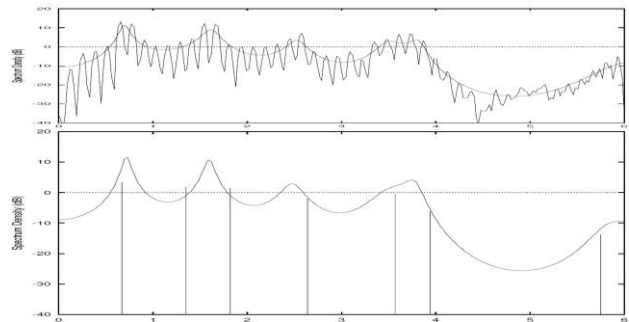


Figure 2. 音声スペクトル、LPC 推定スペクトル、CSM 周波数と強度の例

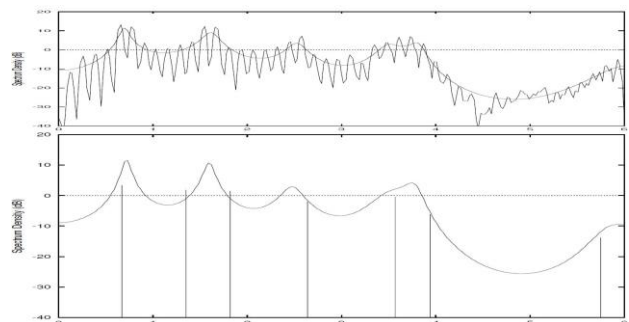


Figure 3. 音声波形と CSM 周波数時間軌跡の例

今後の課題としては、雑音処理の向上、ユーザーへの長時間発声による負担軽減、直観性や GUI の向上などを検討したい。

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] 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.
- [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後援論文集, 5Z-02, Mar. 2016.