

内視鏡を用いた舌操作型ポインティングデバイス

小川 陽子^{1,a)} 島田 伸敬^{2,b)}

1. まえがき

脊髄の損傷による運動障害や先天性四肢機能障害を持つ人々の生活を支援するための一つ的手段として、手を使わないコンピュータの入力インタフェースに期待が持たれており、音声 [1], 眼球運動 [2], [3], 下顎運動 [4] などを用いた方法が提案されている。音声を用いた方法は、古くから研究されており一般に利用されているが、人工呼吸器使用者には発話が困難なため適さない。眼球運動を用いた方法も盛んに研究されているが、重度の障害を持つユーザにとって注視することは心身の負担が大きく、不随意運動による外乱も多い。下顎運動は、外乱は少ないが高位の脊髄損傷者には適用が困難な場合がある。

そこで、我々は重度の四肢機能障害においても運動機能が残りやすく随意的に動かしやすい舌に着目した。舌運動を利用したインタフェースとして、舌マウス [5] や、舌の口腔内ジョイスティック [6] などが提案されている。しかし、舌マウスは、上顎に設置した上下左右の 4 スイッチを舌先上方向に押す必要があり、障害の程度や癖によって十分な操作が行えない場合がある。口腔内ジョイスティックは、啞えたジョイスティックを舌先の小さな力で操作し、その方向と大きさをスティック裏面に取り付けられた特殊な触覚センサで同時に検知する仕組みになっており、特殊なセンサやパーツ加工の必要がありコストがかかる。また、ユーザの障害の程度や癖など個人差に自動的に適応する取り組みはされていない。他にも、下顎底部の 9 点の筋電位差から嚙下、あくび、開口動作の不随意運動と上顎の 3 点 (左, 中, 右) に舌を押し付ける動作を判別する手法が提案されているが [7], こちらも 5 通りの動作の判別に留まっている。

我々は、安価に入手可能な USB 内視鏡を用いて舌の動きを画像シーケンスとして観測し、連続する 2 フレーム間の舌の移動量を回帰することで、舌で扱う相対座標の入力方式を提案する。半透明のシリコンカバーを内視鏡に被

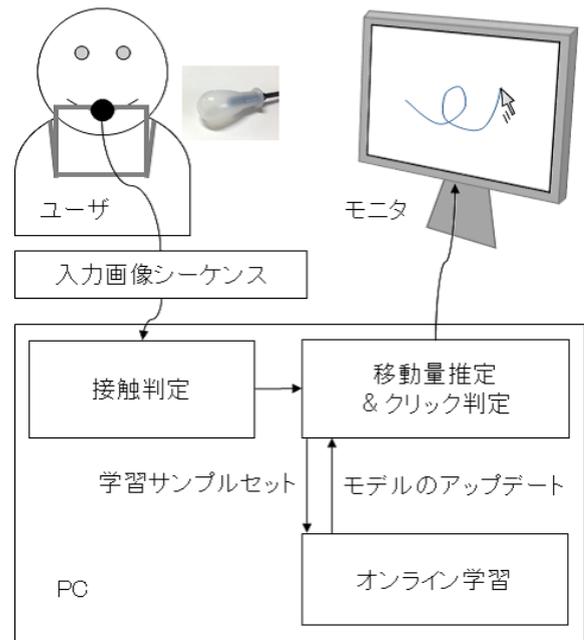


図 1 システム構成

せ、その表面に舌が接して移動するときのみ舌の動きを推定しマウスポインタを動かすことでユーザへの触覚フィードバックと接触の認識の容易化を図る。さらに、操作中に得られた入力画像とクリック位置をもとに転移学習を行い適宜モデルを差し替えることで、操作から得られる個人の特性への適応を試みる。

2. システムの構成

ユーザが自由にコンピュータを操作するためには最低限マウスポインタの移動とクリックができる必要があると考え、本システムはマウスポインタの移動とクリックを可能とする。マウスポインタを動かす際には、動かしたい方向に内視鏡カバー表面を舌でなぞる。クリックをする際は、内視鏡カバー表面に舌を当てたまま一定時間静止する。システム側は、まず内視鏡から画像を取得し、舌がカバー面に触れているか否かを判定する。触れていると判定した場合はマウスポインタの移動量を推定し、実際にポインタを動かす。一定時間舌が接触している場合、クリック判定を行う。クリックが発生すると学習サンプルセットを生成し、一定量以上溜まったら、別スレッドで学習を開始する。

¹ 立命館大学大学院 情報理工学研究科, 〒 525-8577 滋賀県草津市野路東 1 丁目 1-1

² 立命館大学 情報理工学部, 同上

a) ogawa@i.ci.ritsumei.ac.jp

b) shimada@ci.ritsumei.ac.jp

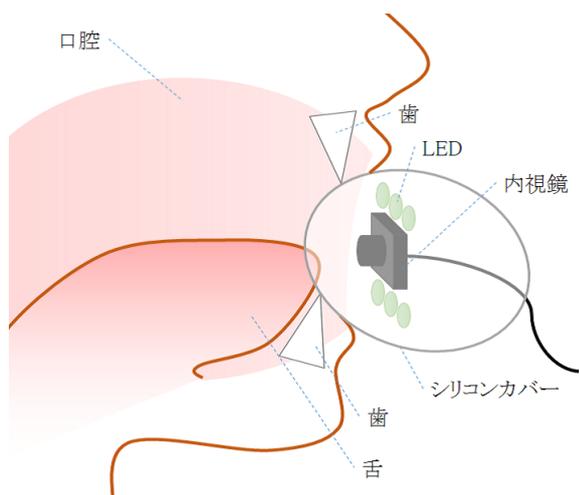


図 2 内視鏡カメラによる舌画像の観測

学習が終わったらモデルをアップデートする。本システムの構成を図 1 に示す。内視鏡は LED 内蔵のものを使用し、図 2 のように舌尖で軽くカバー表面に触れられる位置に固定する。

3. 舌とカバーの接触判定

ユーザ側がコンピュータに何らかの入力を行ったときのフィードバック方法はインタフェースによって異なるが、一般に、画像による観測を用いたインタフェースは非接触であることを利点とし、視覚や聴覚のフィードバックを返すことがほとんどである。しかし、接触を避けるべき用途でなければ、入力を行う器官そのものに触覚を直接提示することが望ましい。そこで、本研究では半透明のシリコンカバーを内視鏡に被せ内視鏡付属の LED を点灯させ、舌がカバー面に接触した際にのみ内視鏡画像に口腔内色の物体が映り込むようにすることで、タッチ感のある入力を可能とする。ユーザは舌がカバーに接触したときにシステム側から反応があると予測でき、システムはユーザの接触が画像の変化から容易に推定できる。

舌とカバーの接触の推定は 2 クラス分類問題に相当する。入力画像と接触の有無の関係を事前に学習することで新規入力画像について接触の有無を判定する。舌とカバーの接触判定には、5 層の畳み込み層と 3 層のプーリング層、3 層の線形層を持つ深層学習モデルである AlexNet[8] をベースに出力層を 2 ノードに変更したものを用いる。入力画像は、縦幅と横幅の小さい方の正方形で画像中心から切り出し、227x227 にリサイズする。以降では、舌と内視鏡カバーの接触がありと判定されたフレームを接触フレームと呼称する。

4. マウスポインタ移動量の推定

ユーザの意図した通りにマウスポインタを動かすためには、時系列のフレーム間で舌がどれだけどちらにシリコンカバー上を動いたかを推定する必要がある。本研究では画

像上での舌位置の座標変位をマウスポインタの移動量として時系列の連続した 2 枚の画像とマウスポインタの移動量の関係を回帰することで、マウスポインタの移動量を推定する。

マウスポインタ移動量の推定には出力層を 2 ノードに変更した AlexNet を拡張して用いる。入力層を 6 チャンネルとし、勾配消失対策として各畳み込み層の後にバッチ正規化 [9] を入れる。また、ここでは画像の分類ではなくマウスポインタの移動量を推定するため、損失関数を平均二乗誤差とすることで CNN を回帰モデルとして扱う。教師データを時系列に連続する接触フレームの相対移動量とし、入力画像の 1~3 チャンネルをフレーム $t-1$ の RGB チャンネル、4~6 チャンネルをフレーム t の RGB チャンネルとする。

5. クリック判定

クリックを実現するためには、[6] のように圧力センサで咬合力を検知する方法も考えられるが、障害の度合いが高い場合、咬合力が十分に出せない場合がある。また、安全面から回路の一部を口腔内に挿入することは可能な限り避けたい。そこで、クリックの判定も内視鏡画像を用いて行う。

内視鏡カバーの下部に舌に触れさせたり離したりといった動作はほとんど力が必要なく簡単であるため、内視鏡カバーに触れる程度に舌を前に出して一定時間静止した場合にクリックを発生させる。このとき、静止条件を推定移動量から求めると、誤検知や検知漏れが発生しうるため、接触の有無のみ 3 節の方法で判別した後シンプルな画像処理で判定する。静止している場合、フレーム間差分が全体的に小さくなり、差分値そのもののばらつきも小さくなると考えられる。そのため、時刻 t において舌と内視鏡カバーの接触しているフレームが $N+1$ フレーム続き、さらに直前 $N+1$ フレームの隣接するフレーム間差分の画素値(絶対値)の平均と標準偏差が閾値を下回った場合を静止と判定する。

6. マウスポインタ移動量のオンライン転移学習

舌を動かす速さや正確さ、可動域、癖などユーザによって舌の動かし方は異なる。ユーザが使用する中で、システムが自動的にその特性を学習し適応すれば、ユーザや周囲の介助者の負担を軽減できる可能性がある。そこで、本システムはユーザの使用履歴に基づきオンラインで転移学習を行う。

学習を行うためには、ある入力画像シーケンスがあったときにユーザがどう動かそうとしたかという意図そのものが教師データとして必要になる。マウスポインタの軌跡のみからユーザの意図を読み取ることは難しいが、クリック

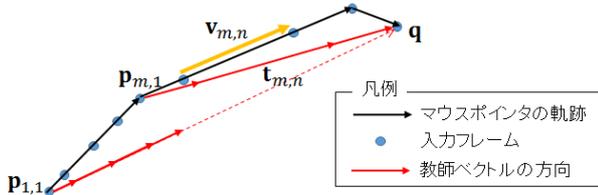


図 3 オンライン転移学習用データの作成

はユーザが意図した場合に起こる動作であると仮定してこれを利用する。ユーザがある位置をクリックする際、まずマウスポインタをその方向へ動かし、ポインタ位置のずれを修正してクリックする。そのため、クリックが行われたフレームでシーケンスを区切り時系列に遡ることで、クリック位置を目標位置とした動作を括り出すことができると考えられる。ここでは、クリックからクリックまでのマウスポインタの移動を1つのモーションとして、その入力画像に対し教師データを生成する方法について述べる。

1つのモーションは1回以上の連続する接触フレーム(以下サブモーション)で構成される。ユーザがサブモーションの最初の入力を行う前にはマウスポインタは停止しており、ユーザはその位置とクリックを行う位置(目標位置)の相対的な位置関係を確認して入力を行う。そのため、サブモーション内の各マウスポインタ移動ベクトルをサブモーションが始まった位置から目標位置の方向に回転させて教師信号とすることがまず考えられる。しかし、単純に回転させるだけでは小さいベクトルを教師信号とする学習サンプルが多くなりやすく、小さい動きに過剰にフィットする。そのため、モーション全体を見てスケールする。モーション全体の中で最も大きな移動ベクトルを含むサブモーションに着目し、そのサブモーションで目標位置に到達するようにスケールを決める。いま、ユーザが点 $p_{1,1}$ からマウスポインタを動かし点 q をクリックしたとする。その間 M 回のサブモーションが発生し、各サブモーションは $N_i = \{N_1, \dots, N_M\}$ フレームを含むとする。さらに、それぞれのフレームを $I_{i,j}$ 、マウスポインタ位置を $p_{i,j}$ 、推定されたマウスポインタの移動ベクトルを $v_{i,k}$ とおく。ただし、それぞれ $(1 \leq i \leq M, 1 \leq j \leq N_i, 2 \leq k \leq N_i)$ である。 $I_{i,k-1}$ から $I_{i,k}$ の教師ベクトル $t_{i,k}$ は、

$$m, n = \arg \max_{i, k} v_{i, k}^T v_{i, k} \quad (1)$$

$$s = \frac{v_{m, n}}{\sum_{k=2}^{N_m} v_{m, k}} \quad (2)$$

$$t_{i, k} = s(q - p_{i, 1}) \quad (3)$$

として求める(図 3 参照)。学習セット $L \in \{I_{i, k-1}, I_{i, k}, t_{i, k}\} (1 \leq i \leq M, 2 \leq k \leq N_i)$ を使ってモデルを更新する。

7. 実験

接触判定と移動量推定の学習には、一人が画面上目一杯

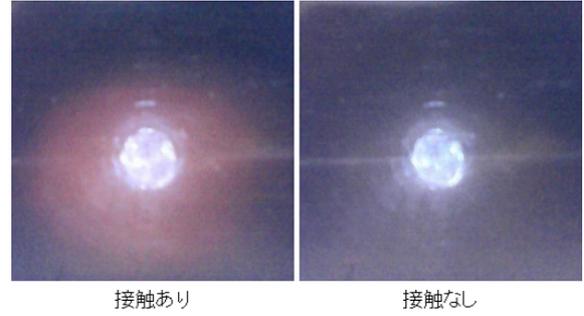


図 4 入力画像例

表 1 実験機スペック

OS	Windows7(64bit)
CPU	Intel®Core™i7-3930K @ 3.20GHz
Memory	32.0GB
GPU	GeForce GTX 780

万遍なく舌を動かした際の画像シーケンスに手動で接触有無のラベルを付けたデータを用いた。入力画像の例を図 4 に示す。接触判定の学習画像は 2644 枚(ポジティブ画像 1303 枚, ネガティブ画像 1341 枚), 検証用画像は 872 枚(ポジティブ画像 432 枚, ネガティブ画像 440 枚)であった。学習性能は 100%, 汎化性能は 99.5%であった。移動量推定には、学習用サンプルとして画像ペア 3782 組, 検証用サンプルとして 1259 組を使用した。学習画像での残差の標準偏差は 35.3pixel, 検証用サンプルでは 48.5pixel であった。

評価基準は、ボタンクリックに要する時間とマウスポインタの軌跡とする。直径 128pixel の円形のボタンを画面中央に 1 つ配置し、それをマウスポインタ初期位置とし、マウスポインタ初期位置とボタン中心の距離が 256pixel になるよう同心円上 8 方向に同サイズの円形ボタンを配置し、被験者には 8 個のボタンを 1 つずつランダムに指示しそれをクリックさせる。ボタンを 1 つクリックするごとにマウスポインタは自動的に中心に移動し、まだクリックしていないボタンが再度ランダムに指示される。転移学習のタイミングを被験者ごとに揃えるため、本実験においては 8 点のクリックが終わる都度学習を行い、学習が終了するまで入力を行わないよう指示した。被験者は、初期モデル、転移学習 1 回、2 回、3 回のモデルでそれぞれ 8 点クリックした後、再度初期モデルを用いて同タスクを行った。処理時間は表 1 に示すスペックの実験機で、接触のないフレームで 25fps, 接触のあるフレームで 8fps ほどであった。クリック判定の接触継続長は 6 フレーム(約 0.8 秒)とした。健常な被験者での結果を図 5 に示す。マウスでは平均 1 秒弱であったが、本デバイスを使った場合、初期モデルで 5 秒ほど、転移学習を行うことで 4 秒ほどであった。クリック判定の時間が 1 秒近くあるため、実質 3, 4 秒ほどでマウスポインタの移動は完了している。転移学習を行った場合の効果はほとんど見られなかった。

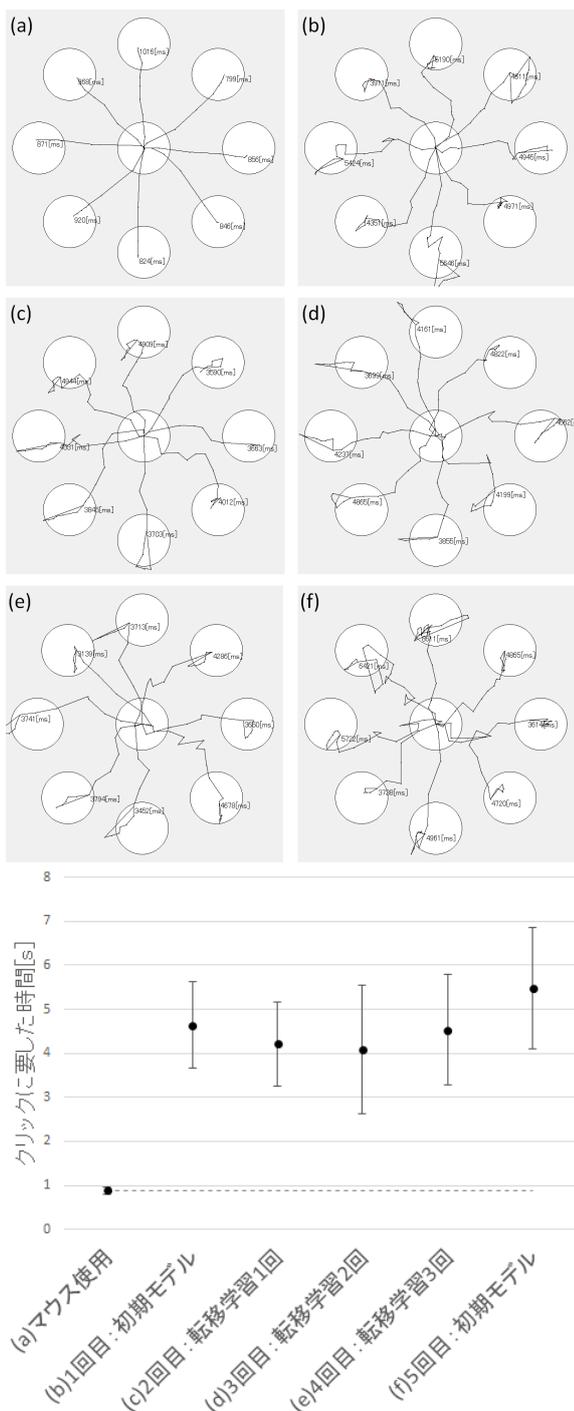


図 5 結果

8. おわりに

本稿では、USB 内視鏡と半透明なシリコンカバーを用いて舌の動きを画像シーケンスとして観測し、舌とシリコンカバーの接触を判定、連続する 2 フレーム間の舌の移動量をマウスポインタの移動量として回帰することで、舌で扱う相対座標の入力方式を提案した。半透明なシリコンカバーを用いることで舌とシリコンカバーの接触判定を容易化し、またユーザへ触覚フィードバックを実現した。さらに、操作中に得られた入力画像とクリック位置から転移学

習を行い、適宜モデルを差し替えることで個人の特性への適応を試行した。実験から、ある程度の慣れは必要なのものの、4~5 秒ほどで 256pixel 離れたボタンのクリックが可能であり、処理時間においても実時間で稼働が可能であることを示した。今回、転移学習の教師データはクリック位置を参照して生成したが、実際はユーザがシステムの反応（ユーザ自身の操作に対するマウスポインタの動き方）に応じて操作を調整するということが起こる。そのため今後は、ユーザ側の調整メカニズムを分析、モデル化し、システム側の調整がユーザにとって使い易くなるように転移学習の部分を改善する必要がある。

参考文献

- [1] L.J.Stifelman , et al. “VoiceNotes: a speech interface for a hand-held voice notetaker,” CHI '93 Proc. of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems, pp.179-186, Amsterdam, The Netherlands, April 24-29, 1993.
- [2] LaCourse, John R., and Francis C. Hludik. “An eye movement communication-control system for the disabled,” IEEE Transactions on Biomedical Eng., vol.37, no.12, pp.1215-1220, Dec. 1990.
- [3] 阿部清彦, 大内幹夫, 大井尚一, 大山実, “画像解析による強膜反射法を用いた視線入力システム,” 映像情報メディア学会誌, vol.57, no.10, pp.1354-1360, 2003.
- [4] 新川拓也, 河内了輔, “下顎運動と舌運動を併用したヒューマンインターフェース,” 生体医工学, vol.44, no.1, pp.94-100, 2006.
- [5] 寺島正二郎, 佐藤栄一, 小竹和夫, 植木一範, 佐々木聡, “重度障害者のための口腔内コントローラーの開発,” バイオメカニズム, vol.20, pp.87-98, 2010.
- [6] 寺島正二郎, 酒井淳一, 大平隆広, 村上肇, 佐藤 栄一, 松澤智由貴, 佐々木聡, 植木一範, “口腔内ジョイスティックの開発：舌を利用した重度障害者のための統合型操作支援装置”I-to-AS”の開発に向けて,” ライフサポート, vol.24, no.4, pp.201-207, 2012.
- [7] 佐々木誠, 荒川峻之, 中山淳, 山口昌樹, “舌骨上筋群の協調運動を利用した舌運動推定法,” 生体医工学, vol.50, no.1, pp.31-37, 2012.
- [8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with deep convolutional neural networks,” In Advances in neural information processing systems (NIPS), pp.1097-1105, Dec. 2012.
- [9] S Ioffe, and C Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” arXiv preprint arXiv:1502.03167, 2015.