

i文鎮: 書くことを楽しくする文鎮

牟田 将史^{1,a)} 石川 優^{1,b)} 里井 大輝^{1,c)} 星野 准一^{1,d)}

概要: 本稿では、人が筆記を行うときに「トメ」「ハネ」「ハライ」の終筆動作がなされたことを筆記音からリアルタイムに識別し、それぞれの終筆動作に対応した効果音を再生する文鎮型デバイス「i文鎮」を提案する。筆記とほぼ同時に効果音を鳴らすことで、手書きによる筆記を楽しむ。手書きによる履歴書の記入や、漢字の書き取り練習のように、大量の文字を手書きするときの苦痛を軽減する効果が期待できる。また終筆動作の識別性能について評価実験を行い、高い割合で正しく識別できることを確認した。

キーワード: 文鎮, 手書き, 音響解析, 機械学習

iBunchin: The Paperweight Makes Your Handwriting Fun!

MUTA MASAFUMI^{1,a)} ISHIKAWA YU^{1,b)} SATOI DAIKI^{1,c)} HOSHINO JUNICHI^{1,d)}

Abstract: We propose the paperweight-like device “iBunchin”. When you take a note, iBunchin identifies which the end of a brush-stroke (“stop”, “hook” or “fade”) performed in real-time, and play the sound effect corresponding to each end of a brush-stroke. By using the sound effect is heard almost simultaneously with the writing, it made writing by hand fun. It is expected that reducing the pain when you write a large number of characters, for example, filling out a resume by hand, or practicing dictation of Chinese characters. We confirmed that iBunchin can identify the end of brush-strokes correctly at a high rate through the efficiency evaluation.

Keywords: paperweight, handwriting, acoustic analysis, machine learning

1. はじめに

紙にペンで文字を書くことは、ワープロやプリンタが普及した現在でも、依然として身近かつ重要な行為である。

手書きは子どもの発達にとって重要であると言われており [1], 子どもが漢字を学習するときには、お手本を指でなぞる練習よりも、お手本を見ながら漢字を書き写す練習の方が学習効果が高いことが報告されている [2]。また、手書きはキーボード入力におけるカナ変換作業が不要であるため、キーボード入力よりも手書きの方が入力速度や認知的負荷の面で優れているとも言われている [3]。しかし、手書きで集中して筆記作業を行っている、手の疲れや作業

の単調さから苦痛を生じるために、筆記を続けるモチベーションや作業効率を保ちにくいという問題がある。この問題に対して、漢字学習についてはゲーム要素を加えることで楽しく書き取り練習を続けられることを図ったアプリケーションが数多く販売されているが、手書きの作業そのものを楽しむには至っておらず、利用できる状況も限られている。

人の動作に音を重畳提示することで楽しさを演出するシステムは、これまで「食べテルミン」[6] や「拡張のどごし感」[7] などが提案されているが、手書きによる筆記についてはあまり扱われてこなかった。

一方で、筆記作業において筆記音を強調してヘッドホンやスピーカーで聞かせると、聞かせなかった場合や無音状態よりも作業効率が高くなることが知られている [4]。また、Andersen and Zhai は、ペンジェスチャー操作に効果音を加えることにより、筆記の正確さや速さなどに差が生じ

¹ 筑波大学
University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan
^{a)} mecab@misosi.ru
^{b)} ishikawa.yu@entcomp.esys.tsukuba.ac.jp
^{c)} satoi.daiki@entcomp.esys.tsukuba.ac.jp
^{d)} jhoshino@esys.tsukuba.ac.jp

ることを示している [5]。このように、紙にペンで筆記を行うときに、音を用いたフィードバックをリアルタイムに与えることで、作業効率や学習効果を向上させることが示唆されている。ところが、楽しさの観点から音のフィードバックを与える試みは十分になされていない。

そこで本稿では、人が文字を書くときに頻繁に行う「トメ」「ハネ」「ハライ」（これらを終筆動作と呼ぶ）に合わせて効果音を再生する文鎮型デバイス「i 文鎮」を提案する。

まず、i 文鎮の内蔵マイクで筆記音を録音し、逐次 PC へ送信する。PC では、受信した音声をリアルタイムに解析する。一定の時間ブロックごとに MFCC と Δ MFCC を連結した特徴量を計算し、非線形 SVM を用いて「トメ」「ハネ」「ハライ」のうちいずれかの終筆動作が行われたか、あるいは行われなかったかを判定する。いずれかの終筆動作が行われたと判定されると、i 文鎮の内蔵スピーカーから、それぞれの終筆動作に対応した効果音を再生する。

i 文鎮では、筆記とほぼ同時にテンポよく効果音を鳴らすことで、手書きによる筆記を楽しむことを図る。本物の文鎮のように紙の上に置くだけで利用するため、紙やペン、机などに細工を施す必要がなく、手軽に利用できる特長がある。

手書きによる履歴書の記入や、漢字の書き取り練習のように、大量の文字を手書きするときの苦痛を軽減する効果が期待できる。特に、漢字の書き取り練習においては、「トメ」「ハネ」「ハライ」の終筆動作を音のフィードバックによって意識しながら練習できるようになるため、学習効果の向上も期待できる。

なお、i 文鎮は「Gugen2013」[8]において「秋月賞」を受賞した。

2. システム構成

2.1 概要

i 文鎮は、幅 15.5cm、高さ 3cm（取っ手を除く）、奥行き 4.3cm の文鎮型デバイスである。i 文鎮の外観を図 1 に示す。電池（単 4 電池 3 本）やマイコン（mbed LPC1768）、Wi-Fi 通信モジュール（GainSpan Wi-Fi module GS1011）、コンデンサマイク 2 台、スピーカーが内蔵されており、図 2 のように簡単に開いてアクセスできる。また、筆記音を入力してから効果音が出されるまでの大まかな処理の流れを図 3 に示す。

2.2 音声入力

i 文鎮の筐体底面からコンデンサマイクを用いて筆記音を取得する。この際、同時に筐体上面に配置されたマイクからも同時に音声を取得し、ノイズキャンセリングを行う。これにより環境音や雑音の影響を小さくし、誤認識を起りにくくした。ノイズキャンセリングは各マイクからの入力を A/D 変換した後にマイコン内で値の差分を取る



図 1 i 文鎮の外観

Fig. 1 Appearance of iBunchin.

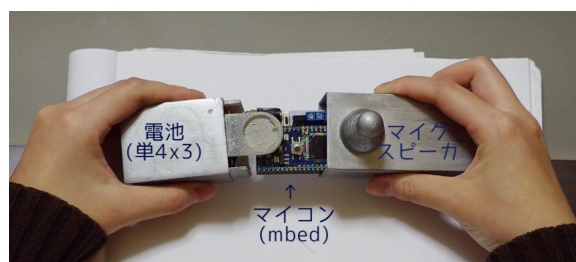


図 2 i 文鎮の内部

Fig. 2 Inside of iBunchin.

ことで行った。A/D 変換器は mbed 搭載のものを利用し、12bit/16kHz で取得した。

2.3 i 文鎮 ↔ PC 間の無線通信

i 文鎮と PC は、あらかじめ Wi-Fi（IEEE 802.11b）を用いたリアルタイム通信を行えるようにしておく。通信プロトコルには UDP を用いた。主な送受信データが音声データであることから、高速な通信が必要となる一方で、多少のデータ誤りは問題にならないためである。

2.4 筆記音のリアルタイム解析

PC で受信した筆記音に対して処理を行い、「トメ」「ハネ」「ハライ」の終筆動作が行われたかどうかを検出する。検出した場合は、どの終筆処理が行われたかを表す信号を Wi-Fi で i 文鎮へ送信する。

2.5 音声出力

終筆処理が行われたことを表す信号を受信した i 文鎮は、MicroSD 内に格納された、対応する効果音の MP3 ファイルを再生する。

3. 筆記音の認識手法

本章では、筆記音から、「トメ」「ハネ」「ハライ」の行為を検出する方法について説明する。

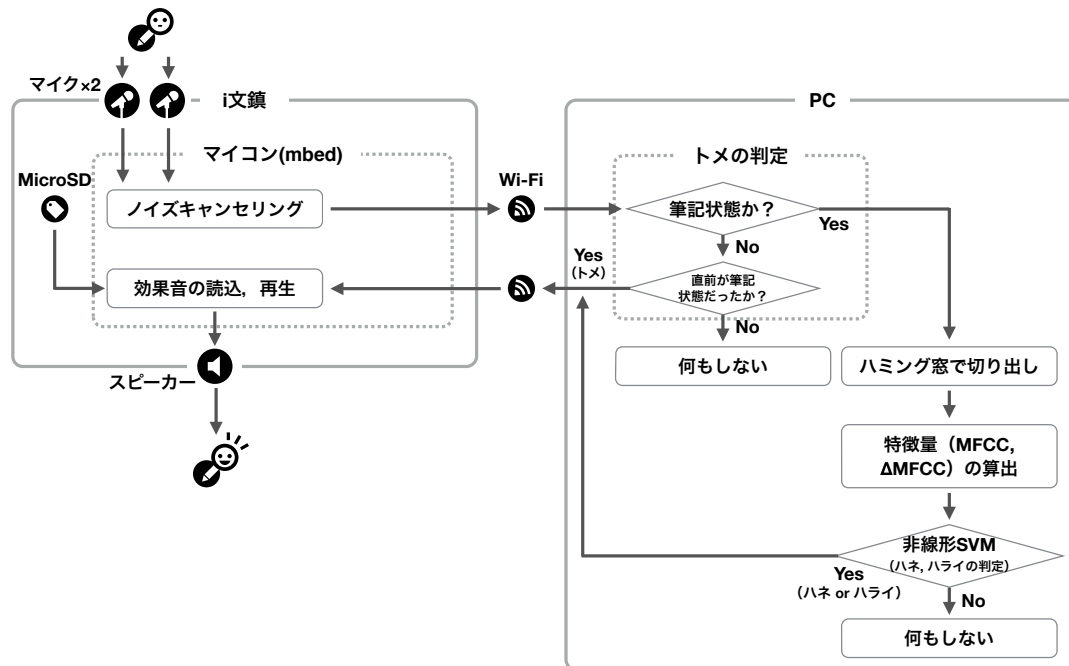


図3 処理の流れ

Fig. 3 Flowchart of iBunchin.

3.1 「筆記状態」「非筆記状態」の検出

まず、現在筆記が行われているか、すなわち筆記具の先が紙面に接触し、かつ移動しているかどうかを判定する。以後、筆記が行われている状態を「筆記状態」行われていない状態を「非筆記状態」という。現在から過去までの一定時間の振幅の平均が閾値を超えている場合は筆記状態とみなし、閾値を超えていない場合は非筆記状態とみなす。

3.2 「トメ」の検出

状態が筆記状態から非筆記状態に切り替わった瞬間に、「トメ」の終筆動作がなされたとみなす。

3.3 「ハネ」「ハライ」の検出

現在が筆記状態である場合、「ハネ」または「ハライ」が行われたかどうかを検出する。これは取得された音声を「ハネ」「ハライ」「その他の筆記音」の3クラスに分類する問題とみなし、機械学習を用いて識別する。取得した音声を256サンプルごとのフレームに分割し、ハミング窓をかけたフレーム単位で処理を行う。

3.3.1 特徴の抽出

まず、フレームから特徴を抽出する。特徴量としてメル周波数ケプストラム係数 (MFCC) 及び MFCC の時間変化である δ MFCC を連結したものをを用いる。MFCC はヒトの周波数知覚特性を考慮した重み付けがなされた特徴量で、音声認識において多く利用される [9]。 δ MFCC は、前後 N フレームの MFCC に対する回帰係数であり、特徴量の時間変化であるデルタ特徴量の 1 つである。MFCC に加え δ 特

徴量を利用することで、認識の精度が向上することが言われており、MFCC と同時に利用されることが多い [10]。

本稿では、MFCC は低次から 17 次元の成分を利用した。また、 δ MFCC は最小二乗法による直線回帰を行うことで求めた。また、求める際に参照する前後のフレーム数 $N=2$ とし、特徴量を求めるフレームも含めた 5 フレームに対して回帰係数を求めた。以上より、1 つのフレームに対して計算される特徴量は MFCC17 次元と δ MFCC17 次元を並べた 34 次元のベクトルとなる。

3.3.2 SVM による分類

得られた特徴量を Support Vector Machine(SVM) を用いて「ハネ」「ハライ」「その他」の 3 クラスに分類する。SVM は教師あり学習を用いる識別手法である [12]。教師データについて、最も他のクラスと近い場所に位置するサンプルとのユークリッド距離を最大化するような超平面を設定することで未知のデータがどのクラスに分類されるかを推定することが出来る。カーネル関数を利用することで非線形な分離が可能となり、精度が向上する。本稿では多項式カーネルを用いた非線形 SVM を利用した。

SVM は本来 2 クラスの分類を行うための手法であるが、今回 3 クラスの分類を行うため、one-against-one のアプローチによる多クラス分類を行う [11]。これは、クラス数 N_c としたとき、 $N_c * (N_c - 1) / 2$ 個の分類器を構築し、すべての 2 クラスの組み合わせについて分類を行い、結果を統合するものである。SVM の実装は scikits.learn を利用した。

教師データは筆者らが実際に「ハネ」「ハライ」「その他」の筆記を行って作成した。筆記音を i 文鎮を用いて録音し、

当該箇所を切り出した。「ハライ」592 フレーム、「ハネ」657 フレーム、「その他」622 フレームの教師データを作成した。

評価実験のためのテストデータとして 1/4 を予約し、3/4 のデータを用いて SVM のパラメータをチューニングし、カーネルの次元を 2、ペナルティ項（誤分類に対する罰則の大きさ）C を 100000 とした。

SVM によってフレームから得られた特徴量が「ハネ」または「ハライ」のクラスに分類された場合、対応する筆記が行われたとみなす。この時、より信頼性を持たせるために、3 フレームの間連続して検出された場合に検出を確定するものとした。

3.4 不感時間の設定

基本的な認識のアルゴリズムは以上だが、この手法では過剰な認識の判定が行われてします。例えば、「ハネ」と「ハライ」の認識は 256 サンプル、すなわち 0.016 秒単位で行われるため、当該終筆動作が行なわれた際に複数回の検出がなされてしまう。また、「ハネ」や「ハライ」の後は通常、非筆記状態に移行するため、直後に「トメ」の検出がなされてしまう。

この問題を防ぐため、検出がされた直後に 0.2 秒の不感時間を設定し、この時間はすべての検出を無効とする。

4. 評価実験

4.1 「ハライ」・「ハネ」の SVM による識別実験

表 1 SVM による識別実験結果

Table 1 Result of the experiment for identification.

	適合率	再現率
ハライ	0.819	0.858
ハネ	0.742	0.733
その他	0.894	0.865

「ハライ」「ハネ」「その他」の筆記音を機械学習により正しく識別できるかどうかを実験により検証した。3.3.2 項で実験のために予約しておいた 1/4 のテストデータに対して識別を行い、各クラスに対して適合率、再現率を測定した。適合率はクラス C であると識別したデータが実際にクラス C であった割合を示し、再現率はクラス C であるデータのうち、クラス C であると予測された割合である。

結果を表 1 に示す。

4.2 実際に利用した際の主観評価実験

筆者らが実際に i 文鎮をおいた状態で文字を書き、正しく動作しているかを主観的に評価した。「トメ」に関しては、概ね正しく検出しているように思われた。また、検出の遅れを感じることもなかった。「ハライ」及び「ハネ」に

ついては検出は行うものの、取り違えて検出（「ハライ」をしたときに「ハネが」発生する。あるいはその逆も）することがあった。また、「あ」の左下のようなカーブを描く際にハライが誤って検出されることがしばしばあった。

5. 考察

4.1 節については、適合率・再現率ともに「ハライ」と「その他」については 8 割を超え、「ハネ」については 7 割を超えた結果となった。これらの結果から概ね正しく識別できていることが分かった。「ハネ」の結果が落ち込んだ理由は、「ハライ」と誤って検出することが多かったためだった。「ハネ」の最後の方のストロークは局所的にハライと似ているため、正しく識別出来なかったのでは無いと思われる。この点に関しては、「ハネ」の教師データとして切り出す部分を検討し直したい。

4.2 節については、誤検出がしばしば起こったものの、検出の遅れもなく概ね期待通り動作した。誤検出については、SVM による終筆動作の識別が何フレーム続いた時に検出を確定するかといった閾値や、不感時間の長さといった後処理にあたる部分の調整である程度抑制出来そうだと思う。また、実際に使ってみる中で、誤検出が必ずしもストレスになるわけではなく、逆に一種の楽しみになるのではないかと感じた。

6. おわりに

本稿では、人が筆記を行うときに「トメ」「ハネ」「ハライ」の終筆動作がなされたことを筆記音からリアルタイムに識別し、それぞれの終筆動作に対応した効果音を再生する文鎮型デバイス「i 文鎮」を提案した。

実験の結果から、「トメ」「ハネ」「ハライ」の終筆動作について、高い割合で正しく識別できることを確認できた。

ただし、誤認識もそれなりの割合で起こっている。今後は、誤認識を抑えるか、あるいは誤認識が起こっても違和感を生まないように改良していくことで実利用をしやすい必要があると考えられる。また、今回は性能評価のみを行ったが、i 文鎮を用いることで文字を書くことを楽しくできるか、筆記作業の効率や認知的負荷を改善できるかを検証していきたい。

加えて、現在の i 文鎮における音の演出は、3 つの終筆動作に対応した効果音を鳴らすだけに留まっている。今後、個人の筆記パターンの特徴や筆記スピードに合わせて鳴らす音を動的に変化させることで、より効果的な楽しみの演出が可能になると考えられる。

なお、現在は実装を容易にするため音声データを PC に転送し音声認識の処理を行っているが、最終的にはすべての処理を i 文鎮内で行えるようにして可搬性を高めたい。

参考文献

- [1] Katya, P.F. and Majnemer, A.: *Handwriting development, competency, and intervention*, Developmental Medicine & Child Neurology, Vol.49, Issue 4, pp.312–317 (2007).
- [2] 小野瀬雅人：幼児・児童におけるなぞり及び視写の練習が書字技能の習得に及ぼす効果，教育心理学研究，Vol.35, No.1, pp.9–16 (1987).
- [3] ムハマド ズルキフリー，田野俊一，岩田満，橋山智訓：日本語のメモ書き作業における手書き入力の有効性，電子情報通信学会論文誌，Vol.J91-D, No.3, pp.771–783 (2008).
- [4] 金 ジョンヒョン，橋田朋子，大谷智子，苗村健：筆記音のフィードバックが単純な筆記作業に及ぼす影響の検討，日本バーチャルリアリティ学会論文誌，Vol.17, No.3, pp.289–292 (2012).
- [5] Andersen, T.H. and Zhai, S.: “Writing with music”: Exploring the use of auditory feedback in gesture interfaces, ACM Transactions on Applied Perception, Vol.7, Issue 3, Article No.17 (2010).
- [6] Kadomura, A., Nakamori, R., Tsukada, K. and Siio, I.: *EaTheremin*, SIGGRAPH Asia 2011, Emerging Technologies, Hong Kong, China, December 13–15 (2011).
- [7] 塚田裕太，金子将大，鹿山晃弘，木下覚：拡張のどごし感，第 17 回日本バーチャルリアリティ学会大会，21E-6 (2012).
- [8] Gugen2013, <http://gugen.jp/>
- [9] Rabiner, L. and Juang, B.H.: *Fundamentals of Speech Recognition*, Prentice Hall PTR, (1993).
- [10] 中川聖一：音声認識研究の動向，電子情報通信学会論文誌，Vol.J83-D-II, No.2, pp.433–457 (2000).
- [11] Knerr, S., Personnaz, L., and Dreyfus, G.: Single-layer learning revisited: a stepwise procedure for building and training a neural network, Neurocomputing NATO ASI Series Vol.68, pp.41–50, (1990).
- [12] Burges, C. J. C.: A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, Vol.2, No.2, pp.121–167 (1998).