

N-best 候補選択と自動形態素分割に基づく スマートウォッチ向け音声入力修正インターフェース

戸羽 遼太郎^{†1} 加藤 恒夫^{†1} 山本 誠一^{†1}

概要：スマートウォッチにおけるテキスト入力は音声認識により短時間に多くの文字入力が可能となるが、細かい修正の必要は常に生じる。本研究では、音声認識で得られる N-best 候補と認識結果の自動形態素分割それぞれの特徴を活かした効率的な修正インターフェースを提案する。N-best 候補選択は認識結果の上位候補に正解がある場合に修正処理を大幅に簡略化し、認識結果の自動形態素分割は修正箇所への素早いアクセスを可能にする。ソフトウェアキーボード入力のみ、音声入力とソフトウェアキーボードを切り替え可能な IME との比較評価を実験協力者 10 名を集めて行った。その結果、提案するインターフェースは入力修正を含めた操作時間が最も短く、使いやすさの主観評価値は最も高く、心理的負荷は最も小さいことが確認された。

1. はじめに

現在、スマートウォッチは内蔵センサを用いた健康管理やスマートフォンからの各種通知の受信など確認用途に多く用いられているが、テキスト入力が簡単になればメッセージ送信や情報検索など用途はさらに広がる。スマートウォッチにおけるテキスト入力は、小さなタッチディスプレイの制約を受けない音声入力が主流とされている。音声入力は多くの文字を短時間で入力できるが、音声認識の誤りや認識後に気づいた入力文中の不備を修正する必要は常に生じる。一方、スマートウォッチの小さなタッチディスプレイ向けのソフトウェアキーボード[1,2,3,4,5,6]や手書き入力[7]なども多く提案され、実用化[8]も進められている。

音声認識の誤りは同音異義語、同音異字の人名、未知語の入力などにおいて形態素単位で生じることが多い。これに対して、修正インターフェースがいくつか提案されている。視線方向を用いる修正インターフェース[9]では、認識結果の予測単語を視線で選択していき、ハンズフリーの入力と修正が可能であることを示した。競合候補を用いた選択形式の修正インターフェース[10]では、選択形式の修正が高い修正性能と操作性を持つことを示したが、一方で未知語の修正が課題として挙げられた。未知語への対策として、単語候補に加え、キーボード入力による修正を可能としたタッチスクリーン型デバイス[11]では、高い操作性に加え、入力文に柔軟に対応できることが示された。しかし、小さなタッチディスプレイのスマートウォッチで認識候補を用いた選択形式の修正を行うには、画面上に表示する認識結果や候補の閲覧性を保ちつつ、修正を行いやすいユーザーインターフェース設計が必要になる。

そこで本研究では、音声認識で得られた N-best 候補と認識結果の自動形態素分割それぞれの特徴を活かした効率的なスマートウォッチ向け音声入力修正インターフェースを

提案する。N-best 候補選択は認識結果の上位候補に正解がある場合に修正処理を大幅に簡略化し、認識結果の自動形態素分割は修正箇所への素早いアクセスを可能にする。また、直観的なレイアウト設計と操作性で画面の狭いスマートウォッチでも簡単に修正を行うことが可能となる。実験協力者 10 名を集めて比較実験を行ったので報告する。

2. スマートウォッチ向け音声入力修正インターフェース

2.1 N-best 候補選択と自動形態素分割に基づく音声入力修正インターフェース

提案方式では、まず音声認識でテキストを入力し、誤認識や要修正箇所が生じた場合にタッチ操作で修正する。音声認識の N-best 候補の選択画面と第 1 位の音声認識結果を自動的に形態素に分割して形態素単位で選択できる形態素分割表示画面により、素早い修正を可能とする。提案インターフェースを構成する音声入力画面、N-best 候補表示画面、形態素分割表示画面、キーボード入力画面の 4 つの画面について述べる。

(1) 音声入力画面

文字入力は図 1 のパネル a) に示す音声入力画面から開始される。音声入力が開始されると 10 秒間音声認識待ち状態となり、その間にスマートウォッチのマイクに向けて発話すると音声認識結果が画面に表示される。音声認識エンジンは Android 標準 API の Speech Recognizer[12]を用いた。

(2) N-best 候補表示画面

修正が必要な場合、まず図 1 のパネル b) に示す N-best 候補表示画面を利用する。この画面では音声認識の N-best 候補 3 文（第 2・3・4 候補）を表示し、候補に所望の文が含まれれば、選択して修正することができる。N-best 候補は大量に得られるが、所望の文である確率の低い候補の確認に時間を取られすぎないように、今回は第 4 候補まで表示する。候補文の閲覧性を高めるために、一候補ずつ別の画面に表示し、左右方向のスワイプ操作によって別の候補の表

^{†1} 同志社大学
Doshisha University



図 1 音声入力修正インターフェース. a) 音声入力画面, b) N-best 候補表示画面, c) 形態素分割表示画面

表 1 形態素分割表示に対するタッチ操作と対応機能

タッチ操作	対応機能
矢印ボタンをタップ	形態素の後に句読点挿入
右フリック	形態素の削除
形態素をタップ	キーボード呼び出し

示に切り替えられるようにした.

(3) 形態素分割表示画面

N-best 候補表示画面の候補に正しい文が無かった場合, 「キーボードで修正」ボタンにより図 1 のパネル c) に示す形態素分割表示画面に遷移する. この画面では音声認識結果を形態素ごとに区切り, 縦方向に列挙する. これにより, 文中の修正箇所まで 1 文字ずつカーソル移動を繰り返す必要がなくなり, 縦方向のSwipeだけで修正箇所に移動できる. 形態素解析処理には goo ラボの形態素解析 API[13] を用いた.

さらに, 文の修正効率を高めるために, 表 1 に示すとおり, 形態素の左にある矢印ボタンにより形態素の後に句読点を挿入, 形態素を右方向にフリックして形態素を削除, 形態素をタップしてキーボードを呼び出し, という 3 つの機能を形態素分割表示に対するタッチ操作として割り当てた.

(4) キーボード入力画面

形態素分割画面で修正対象の形態素をタップするとキーボード入力画面に移行する. 選択した形態素の修正にはソフトウェアキーボードを使用する. ここでは図 2 の左パネルに示す Google 日本語入力を採用した. このキーボードは, スマートフォンと共通のテンキー上でのフリック操作で入力するので, 本稿では以降テンキーフリック式入力と呼ぶ.

2.2 比較実験用インターフェース

インターフェースの比較実験用に以下の 2 種類のインターフェースを用意した.

2.2.1 キーボード入力のみ

図 2 の左パネルに示すテンキーフリック式入力を用いて初めから文を入力してもらう. かな漢字変換や予測変換機能が使用できる.



図 2 左: Google 日本語入力, 右: 音声入力 IME

2.2.2 IME 切り替え

音声入力を行った後, 修正が必要な場合はテンキーフリック式入力で修正する. 音声入力 IME は図 2 の右パネルに示す音声入力 IME を用いる. 音声入力 IME は 2.1 節(1)の音声入力画面に準じているが, 修正を効率的にするために, バックスペースキーを配置し, 入力文中にカーソルを設けた. キーボードあるいは音声どちらか片方だけの入力を避けるために, まずは音声入力を行い, 要修正箇所がある場合にはテンキーフリック式入力に切り替えて, 修正をしてもらう.

3. 評価実験手法

3.1 実験概要

3 種類のインターフェースの比較実験を実施した. 実験協力者には指定した短文をできるだけ間違えないよう入力してもらう. 比較実験後, SUS(System Usability Scale)[14] によりユーザビリティを, NASA-TLX(NASA Task Load Index)[15]により心理的負荷を評価した.

3.2 実験協力者

実験協力者は, 21~23 歳の男性 10 名であった. そのうち, 9 名が右利き, 1 名が左利き, 全員がスマートウォッチの使用経験が無かった. 実験協力者のうち, 7 名が普段からスマートフォンでテンキーフリック式入力を使用していた. 残りのうち 1 名はトグル入力を行っていた. 残り 2 名は QWERTY キーボードを使用していた.

3.3 使用したデバイス

本研究では, HUAWEI WATCH2 を使用した. 解像度は 326[ppi], 画面の直径は約 30.5[mm]である.

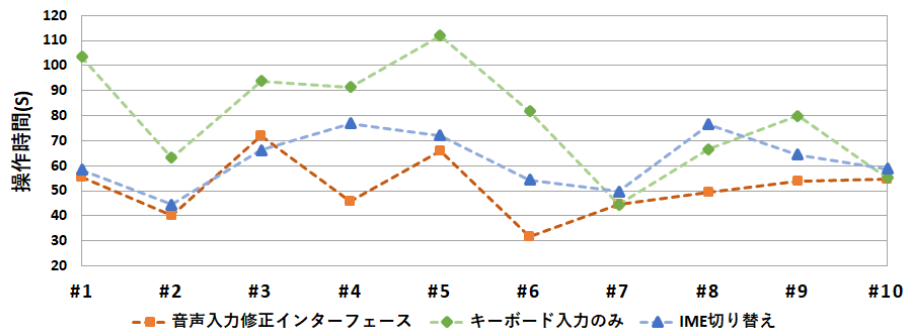


図 3 実験協力者ごとの各入力手法における 1 文入力の平均操作時間(秒)

(入力課題) 阪田先輩は普段は優しいが、怒ると怖い
(認識結果) 坂田先輩は普段は優しいが怒ると怖い

図 4 実験に使用する短文（入力課題文）と認識結果の例

3.4 用意した短文

漢字変換前のかな文字数が約 25 文字の文を 50 文用意し、音声入力修正インターフェース、キーボード入力のみ、IME 切り替えの各手法でそれぞれランダムに選択された 5 文、計 15 文を入力してもらう。順序効果の影響を避けるために、各手法の実施順は実験協力者ごとにランダムに割り当てた。全ての短文は音声入力を用いた場合に、漢字の修正や句読点の挿入が 1 箇所あるいは 2 箇所必要になるよう、事前に音声認識による確認を行っている。図 4 に実験で使った短文の入力課題例と音声認識を行った場合の認識結果例を示す。

3.5 実験手順

実験は音声入力中に雑音の入らない静かな室内で行う。最初に実験協力者は NASA-TLX の評価練習に取り組む。練習用タスクは正負の四則演算問題とする。評価練習後、実験協力者はスマートウォッチ上のキーボード入力と音声入力の操作を 15 分間練習する。操作練習後、音声入力修正インターフェース、キーボード入力のみ、IME 切り替えの 3 手法でそれぞれ 5 文の短文を入力する。また、SUS と NASA-TLX を各手法の入力評価後にその都度行う。

3.6 インターフェースの評価

評価指標として、入力開始から修正を含む 1 文入力が完了するまでの平均操作時間（秒）を用いる。また、SUS と NASA-TLX を用いる。SUS は表 3 に示す 10 項目を 7 段階のアンケート形式で質問し、各入力手法に対する利便性と有用性を計測する。NASA-TLX は各入力手法が作業中に実験協力者に与える心理的負荷を計測する。

4. 評価実験結果

4.1 操作時間

各入力手法における 1 文入力の平均操作時間の結果を図 3 に示す。実験協力者の平均値は音声入力修正インターフェースが 51.4 秒、キーボード入力のみが 79.2 秒、IME

表 2 入力・修正手法による 1 文入力の平均操作時間

入力手法	平均操作時間(秒)
キーボード入力のみ	79.16
IME 切り替え	62.19
音声入力修正インターフェース（形態素分割表示画面からキーボード入力を用いた場合）	59.10
音声入力修正インターフェース（N-best 候補選択のみで修正が完了した場合）	24.80

切り替えが 62.2 秒であった。音声入力修正インターフェースの値は、他入力手法と比べて操作時間が短いことが分かる。各入力手法の平均操作時間に対して分散分析を行った結果、有意水準 0.05 で有意差があることを確認した。音声入力修正インターフェースと IME 切り替えの差が大きい実験協力者（#4, #6, #8, #9）に着目すると、4 名とも N-best 選択により修正を行っている。そこで、音声入力修正インターフェースの N-best 候補選択のみで修正を完了した場合と形態素分割表示画面からキーボード入力を用いた場合を場合分けして、キーボード入力のみ、IME 切り替えと比較した平均操作時間を表 2 に示す。表 2 では N-best 候補選択による修正を含めなかった場合の音声入力修正インターフェースの平均速度は IME 切り替えの場合と 3 秒しか差が無く、N-best 候補選択による修正の効率性が明らかになった。また、図 3 より実験協力者 10 名中 9 名で音声入力修正インターフェースが最も素早く修正できており、短時間に習得できるインターフェースであることが分かった。

4.2 SUS

SUS 評価に使用した項目と各入力手法におけるスコアを表 3 に示す。音声入力修正インターフェースに着目すると、10 項目中 8 項目で最も高い数値となっており、100 点満点に換算した結果を見ても最も高い数値となった。ここで SUS スコアについて Bangor らの評価基準[16]では、90 点以上で Best imaginable, 80 点以上で Excellent, 70 点以上で Good, 50 点以上で OK, 30 点以上で Poor, 30 点未満で Worst imaginable と分類している。音声入力修正インターフェースの得点は OK に相当するが、これはシステムの有用性の

表 3 評価項目と各入力手法における平均スコア

項目	キーボード	音声入力	IME
	入力のみ	修正 IF	切り替え
頻繁に使いたいと思う	2.5	4.5	4.5
複雑だと感じる*	4.3	4.3	4.3
簡単に使える	2.8	4.9	4.9
技術者のサポートが必要*	4.6	3.3	3.3
機能がよくまとまっている	3.0	4.6	4.6
多くの矛盾があると思う*	4.8	5.5	5.5
使い方をすぐに学べる	4.2	4.7	4.7
とても扱いにくい*	2.5	4.4	4.4
使いことに自信がある	3.9	4.9	4.9
使用前に多く学ぶ必要がある*	4.1	3.2	3.2
得点 (100 点満点)	56.2	68.6	59.0

各項目の7段階評価を好評価が高い数値になるよう0から6点に変換した。

またその合計を1.67倍して約100点満点に得点化した。*は逆転項目である。

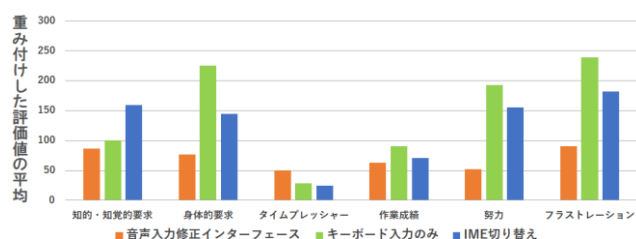


図 5 各入力手法における項目ごとの重み付けした評価値の平均

観点から 使いやすいことを示す最低限のレベルであった。なお、SUS スコアを分散分析したところ、有意水準 0.05 で各入力手法間に有意差は見られなかった。

4.3 NASA-TLX

NASA-TLX の 6 項目の平均スコアを図 5 に示す。評価時、実験協力者には項目を一对比較してもらう。これを重みとして各スコアに掛け合わせたものの総和を、一对比較回数で割った WWL (Weighted Workload) 値を算出した。各入力手法の平均 WWL 値は音声入力修正インターフェースが 27.9、キーボード入力のみが 58.5、IME 切り替えが 49.1 と音声入力修正インターフェースが最も低い値となった。分散分析を行った結果、有意水準 0.05 で有意差が確認でき、音声入力修正インターフェースが他入力手法より入力中に実験協力者に与える心理的負荷が少ないことが分かった。図 5 を見ると「身体的要求」「努力」「フラストレーション」が他手法よりも小さな値となっている。N-best 候補選択による修正や、認識結果を形態素単位に分割して表示し、スワイプ操作で修正箇所を指定する操作法が実験協力者の心理的負荷を和らげたと考えられる。

5. おわりに

本研究では、スマートウォッチ向けに音声入力の結果を効率的に修正できるインターフェースを作成した。N-best 候補選択による修正と素早く修正箇所へアクセスできるよう自動で形態素ごとに認識結果を分割する機能を実装して、小さな操作画面でも音声認識結果の素早い修正が可能となった。

作成したインターフェースの比較評価を実験協力者 10 名に対して行った。比較実験ではキーボード入力のみ、IME 切り替え使用時と比べ、入力に要した時間は最も短くなり、文字入力性能が高いことを確認できた。また、音声入力修正インターフェースは利用者に与えるストレスが最も少なく、実験協力者にとって最も使いやすいインターフェースであることを確認した。

参考文献

- [1] 石井, 箱田, 志築, 田中, 「Flickkey: 超小型タッチパネル端末におけるフリック操作を活用した QWERTY キーボード」, 情報処理学会研究報告, Vol.2015-HCI-164, No.6, pp.1-8, 2015.
- [2] 下岡, 山名, 「スマートウォッチにおけるタップ動作の少ない仮名文字入力手法」, DEIM Forum 2017, I3-2, 2017.
- [3] 秋田, 田中, 佐川, 「スライドインによるスマートウォッチ向けの文字入力手法」, インタラクシオン 2018, pp.276-281, 2018.
- [4] 梅澤, 原, 大澤, 「腕時計型端末における画面占有面積を抑えたかな文字入力手法」, DICO 2017, pp.1746-1753, 2017.
- [5] 東條, 加藤, 山本, 「BubbleSlide: フリック操作の規則性を高めたスマートウォッチ向け日本語かな入力インターフェース」. 情報処理学会 インタラクシオン 2018, pp.1048-1053, 2018.
- [6] T.Tojo, T.Kato, S.Yamaoto, “BubbleFlick: Investigating Effective Interface for Japanese Text Entry”, Proc.MobileHCI, 44:1-12, 2018.
- [7] M.Gordon, T.Ouyang, S.Zhai, “WatchWriter: Tap and Gesture Typing on a Smartwatch Miniature Keyboard with Statistical Decoding”. Proc.CHI2016, pp.3817-3821, 2016.
- [8] “TouchOneKeyboard”, <http://www.touchone.net/>, (参照 2018-12-12).
- [9] K.Vertanen, D.J.C.MacKay, “Speech Dasher: Fast Writing using Speech and Gaze”. Proc.CHI2010, pp.595-598, 2010.
- [10] 緒方, 後藤, 「音声訂正: 選択操作による効率的な誤り訂正が可能な音声入力インターフェース」, 情報処理学会論文誌 Vol.48, No.1, pp.375-385, 2007.
- [11] K.Vertanen, P.O.Kristensson, “Parakeet: A Continuous Speech Recognition for Mobile Touch-Screen Device”, Proc.IUI2009 p.237-246, 2009.
- [12] “Android Speech Recognizer API”, <https://developer.android.com/reference/android/speech/SpeechRecognizer>, (参照 2018-12-04).
- [13] “形態素解析 API goo ラボ”, <https://labs.goo.ne.jp/api/jp/morphological-analysis/>, (参照 2018-12-13).
- [14] J.Brooke. “SUS -A quick and dirty usability scale”, Usability Evaluation in Industry, pp.189-194, 1996.
- [15] S.G.Hart, L.E.Staveland, “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research”, Human mental workload, Vol.1, No.3, pp.139-183, 1988.
- [16] A.Bangor, P.Kortum, J.Miller, “Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale”, Journal of Usability Studies, Vol.4 Issue3, pp.114-123, 2009.