

圧力センサによる点字解読に特化したスペルコレクタ

代表研究者 伊藤 祥一 長野工業高等専門学校・電子情報工学科・准教授

1 はじめに

厚生労働省の調査によると、我が国における視覚障害者は約 31 万人で、そのうち 1 割程度が点字を扱うことができる。視覚障害者の 8 割以上は病気やケガを理由とする中途失明者である。先天的な視覚障害者は幼少の頃から特別支援学校(昔の呼び方では盲学校)に通うことで点字の文法を身につけることができる。若いうちから指先をトレーニングすることで点字の点を素早く読み取れる鋭敏な感覚がやしなわれる。中途失明者は年齢的な問題から、新しい言語として点字を受け入れることが難しく、さらに指先の感覚を鍛えることも難しいため、点字を習得することが難しい。一般に、中学卒業(14 歳)程度を越えると点字の習得がかなり難しくなると言われている。一方で 21 世紀に入り、アクセシビリティやバリアフリーという考え方が社会に浸透してくることに歩調を合わせ、公共の場を中心として各種の案内に点字が併記されることが多くなった。今では家電品にも点字が多く使われている。しかし点字を読めない視覚障害者はこれらの支援を受けることができず、結果として家にこもることが多くなりがちである。自分に代わって文字を読んでくれる介助者がいて外出できたとしても、ATM の残額表示のようなプライベートな情報を読んでもらうことは避けたいであろう。もし、視覚障害者本人に代わって機械が点字を読んでもくれば、視覚障害者の生活品質の向上が期待できる。そこで我々は数年前から、常に身につけておいて点字を読みたいときにサッとなぞれば点字を読み上げてくれる機械、ウェアラブル点字リーダーの開発を進めている。

ウェアラブル点字リーダーの利用者は指先に装着した圧力センサで点字をなぞる。すると圧力分布データが、利用者のポケットの中の小型コンピュータに入力され、点字を翻訳した文章が合成音声によってイヤホンから聞こえる。現時点で使用している既製品の圧力センサは大型で指先に装着することはできないが、それ以外のシステムはほぼ完成している。圧力センサで点字をなぞったとき、指先でおさえて横にすべらせるという動作の原理上、点字の凸点をおさえる力にばらつきが生じ、本来あるべき点が欠落したり、逆に点のない部分に点があるように圧力がかかってしまったりしてしまうことがある。点字の文字列をまっすぐ水平になぞることも難しい。このような圧力データから点の並びを再構成すると当然まちがった点字の並びが取得されることになり、それを翻訳して音声として耳に届けても無意味なものになってしまう。これを改善するためには絶対に読み取りエラーをおこさない圧力センサを開発するか、もしくは、誤りを含んだ圧力データが来ても最終結果を正しく直してしまう仕組みがあればよい。本研究では後者の立場で、圧力センサで点字をなぞったとき固有のエラーを調べ、誤りを含んだ点の配列から正しい翻訳結果に修正するスペルコレクタの仕組みを開発する。読み取りの精度を向上できるだけでなく、より安価で簡単な仕組みの圧力センサでも十分な読み取り精度を実現する道を開くことにもつながる。

2 ウェアラブル点字リーダー

我々が開発しているウェアラブル点字リーダーの全体像を図 1 に示す。図 1 の右上は名刺サイズのコンピュータ(Raspberry Pi 4 Model B)である。図 1 右下のモバイルバッテリー(USB PD)からの給電で動作する。図 1 左の基板は 15cm 角の大きさの圧力センサである。実際に点字を読み取る部分は 1.2cm 角で、センサ基板中央下部にある。ここで読み取った点字の情報がコンピュータに送られ、音声に変換されたものが図 1 下のイヤホンに無線(Bluetooth)で送信される。現時点では圧力センサがまだ大きく、指先に装着できる段階にはないが、開発会社と協力して改良を進めている。今回はこのセンサに点字を押し当てて点字の側を横にスライドさせることで、点字を指先でなぞっている様子を模擬的に再現して実験をおこなった。

28 種類の点字文字列を読み込ませたところ、現行のスペルコレクタを組み込んでいないシステムでは、14 種類は点のパターンを間違えずに読むことができた。28 種類のうち 5 つは点の誤りが 2 つ以内であった。このような微妙な読み取りエラーをスペルコレクタの仕組みを使って改善しようというのが本研究の目的である。残りの 9 つは点字を均等な力で圧力センサに押し当てて水平に動かす事が難しいなどの問題でうまく読み取りができなかった。圧力センサを改良して、センサで点字をなぞることができれば改善されることを期

待している。



図1 開発中のウェアラブル点字リーダーの全体像

3 点字のサイズ

現在、世界各国で使われている点字は横2列×縦3行の6点式がほとんどである。図2は6点式点字の点間距離を定義したものである。図2の a , b , p , q の具体的な値を表1に示す(値に幅があるものは中央値を示した)。点字のサイズは我が国においては日本工業規格(JIS)で定められている。表1には現在使われている点字で採用されている T0923:2009 のものを示した。我が国における点字の規格で最新のものは JIS T0921:2017 となり、JIS T0923:2009 とは微妙にサイズが異なっている(T0921:2017 は ISO 17049:2013 との整合性を優先して T0923:2009 と比べてわずかに大型化している)。

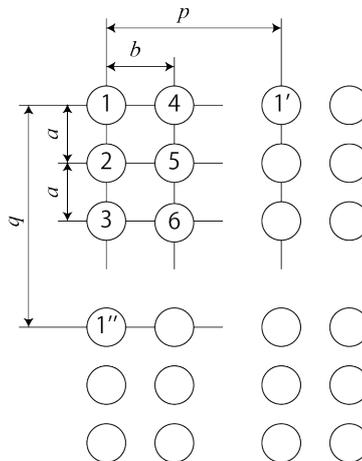


図2 点字の点間距離の定義

表1 点字のサイズ(図2の各変数に対応、単位[mm])

国名	a	b	p	q
日本	2.40	2.25	5.70	13.00
イギリス	2.50	2.50	6.00	10.00
アメリカ	2.34	2.34	6.22	10.16
カナダ	2.34	2.34	6.22	10.16
オーストラリア	2.40	2.40	6.50	10.50

4 点字解読に特化したスペルコレクタ

4-1 点の過不足の訂正

指先に装着した圧力センサで点字をなぞるとい動作の原理上、点字の凸点をおさえる力にばらつきが生じたり、点の並びに対して水平に指を移動することが正確にはできないことから、圧力分布データをコンピュータで分析して点の並びを再構成した際、本来あるべき点が欠落したり、逆に点のない部分に点があるようにみえてしまうことがある。図3は、「トショカン」という点字を圧力センサでなぞったデータを処理した際に、点が1つ欠落し、「ノショカン」という点字として再構成されてしまった例である。スペルコレクタを使用しない場合、「ノショカン」がそのまま読み上げられてしまい、利用者にとっては役に立たない情報になってしまう。

本研究では、スペルコレクタ用の辞書として、IPA辞書(約39万語)、NAIST-jdic(約48万語)、NEologd(約319万語)を統合し、重複を削除した約210万語からなる辞書を作成した。点字は「今日はいいい天気ですね」という文章を「きょーわ いいい てんきですね」のように、読んだままを文字として分かち書きするというルールがあるが、これらの辞書には表記データとあわせて発音データも収録されているため、発音データをそのまま点字表記とみなして使用する。圧力センサで検出された生データを処理して点の並びに再構成したら、辞書内を検索し、点の並びがもっとも近いものを訂正候補として選ぶ。図3の例では、「ノショカン」という単語は辞書に存在せず、点の並びがもっとも近い単語として「トショカン」が訂正候補として選ばれる。

指先で点字をなぞるとい動作から、力のかけ具合が強すぎたり弱すぎたりといった場所は点字の文の全体に均等に散らばるのではなく、特定の場所に偏るであろう事が想定される。たとえば「れろん」という単語は辞書にないので点の並びが近いものを探すと、2点違いで「れもん」あるいは「ろーん」に補正できる。前者は隣り合う2点を補い、後者は1文字分離れた2点を無視することで得られる。点の数だけみれば同じ2点エラーであるが、圧力センサで点字をなぞるとい動作を考えると、おそらく隣接する場所で欠落が起きている前者のほうがエラーとしては起こりえる可能性が高いと考えられるので、「れもん」を訂正候補として選ぶ。

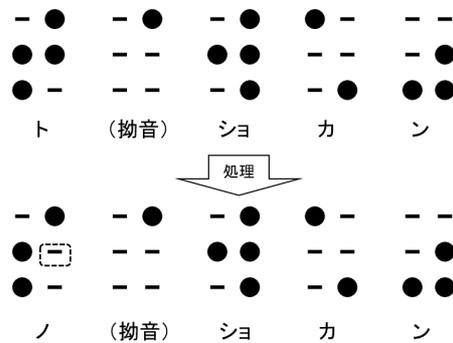


図3 「トショカン」を読み取ったところ点が1つ欠落した例

4-2 逆さ点字の訂正

筒状の容器のフタにあたる部分に点字が印刷されている場合を考える。例えば現在市販されている缶ビールのフタ部分には誤飲防止策として「おさけ」という点字が刻印されている。調味料のソースのフタには「そーす」と刻印されているものもある。しかし点字は単なる点の並びで文字を表しているため、ひらがなのようにどちらが文字として正しい向きかということは一見してもわからない。上下が逆さ向きの点字を左から右に読むと、「おさけ」は「まきを」、「そーす」は「ひーち」とそれぞれ読めてしまい、意味をなさない。我々が開発したスペルコレクタでは、読み取った点の並びを4-1節で説明した手順で辞書検索する際、図4のように上下左右を反転した点の並びでも同時に検索する(ここでは上下左右を反転したものを「逆さ点字」と呼び、反転前のものを「正点字」と呼ぶ)。圧力センサによる読み取りエラーはそれほど多くないという前提で、もともと読み込まれた点の配置に対して過不足が一番少ない単語を訂正候補として返す。

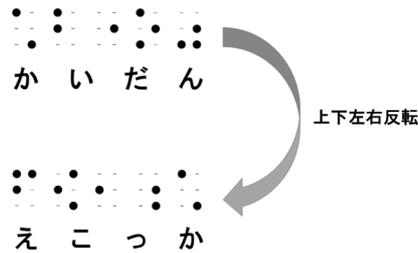


図4 逆さ点字の例

4-3 助詞・助動詞の処理の改善

我々が開発した辞書には自立語のみが収録されているため、自立語に助詞や助動詞が附属した場合はそれらを取り除いてから語の訂正を行う必要がある。まず単純な手法(方法A)として、語の末尾に「に」「を」「は」のような助詞として使われる文字が付随している場合は機械的にそれを取り除いてから辞書を検索し、訂正候補につけなおす、といった処理を実装した。英語の文章は単語ごとにスペースやカンマで区切られているが、日本語の文章は区切りなく一続きになっているため、日本語の文章を解析するためには本来は品詞ごとに切りわけ作業が必要である。しかし点字の文章ははじめから文節ごとにわけて書かれているのでその必要はなく、1つ1つの単語の末尾だけを見ればよい。この方法を試してみると、「エレベータに 乗る」「兄へ手紙を 書く」のような文章を正しくスペル訂正できるようになった。

しかし、方法Aでは「おから」「どんより」など、単語の末尾に「から」「より」のような付属語を含む単語を機械的に切り分けて「訂正」してしまうケースが見られた。これらは独立した単語なので機械的に切り分けてはいけない。そこで日本語形態素解析システムである Mecab(メカブ)を補助的に用いて品詞をある程度考慮した訂正を試みた(方法B)。図5は「寒いので」という文をMecabにより品詞に分解した例である。「寒いので」の「ので」が接続助詞として解析されている。同様に「おから」「どんより」を分析すると、図6と図7のように一つの単語として解析されており、この場合は末尾の「から」「より」を単語から切り分ける必要はない。このような解析結果を用いて助詞・助動詞の識別を行う。

さむい 形容詞,自立,*,*,形容詞・アウオ段,基本形,さむい,サムイ,サムイ
 ので 助詞,接続助詞,*,*,*,*,ので,ノデ,ノデ
 EOS

図5 Mecabにより「寒いので」を形態素解析した結果

おから 名詞,一般,*,*,*,*,おから,オカラ,オカラ
 EOS

図6 Mecabにより「おから」を形態素解析した結果

どんより 副詞,一般,*,*,*,*,どんより,ドンヨリ,ドンヨリ
 EOS

図7 Mecabにより「どんより」を形態素解析した結果

Mecab を用いて助詞と助動詞の判別を採り入れたスペル訂正は以下の手順で行う。

1. 分かち書きされた点字データを日本語に翻訳する
2. 日本語に翻訳したテキストを Mecab によって解析し、品詞 ID(id)と語(surface)を取得する。
3. 品詞 ID が特定の値(13 以上 25 以下、または 32)であれば助詞・助動詞があると判断する。このときの surface が助詞・助動詞である。
4. 翻訳前の点字データの末尾を助詞・助動詞の文字数分だけ削る。
5. 助詞・助動詞を取り除いた点字データが辞書内に存在するか調べ、訂正を行う。
6. 訂正後に手順 3 で削除した助詞・助動詞をつけ足す。

4-4 Mecab による処理を考慮した最終出力選択

Mecab を用いて助詞と助動詞の判別を採り入れた場合、「おから」「どんより」のように、単語の一部として「から」「より」などの助詞・助動詞と同じ文字の並びを持つ単語をうまく扱えるようになった。しかし、別の側面で Mecab を使った場合にうまく処理できないケースがあることも判明した。Mecab は正しい日本語の文章の形態素を解析するソフトウェアであり、誤りを含む単語をうまく区別できない。たとえば「経済学は」という点字の文章をなぞって「けいざいがくは」というデータが入ってくれば Mecab は「経済」と「学」という単語を知っている(「経済学」という名詞は辞書にないようである)のでこの文節を正しく品詞ごとに分解することができる(図 8)。ところが点字の読み取りエラーから「えいざいがくは」というデータが入ってきたとき、Mecab は「えいざいがく」という単語を知らないの、文節末尾の「は」を助詞として見分けることができず、「えい」という名詞と「ざいがくは」という固有名詞に分解してしまう(図 9)。このように、入力データに誤りがあれば直すというスペルコレクタの動作に対して、入力データに誤りがあると Mecab では正しく品詞に分解できないということになってしまう。今回は、まず 4-1 節と 4-2 節で述べた方法で訂正を試みる。正点字の訂正結果と逆点字の訂正結果のうち、入力された正点字の点の並びに近い方を選ぶ。これを A とする。A として正点字が選ばれたら正点字を、逆点字が選ばれたら逆点字を、それぞれ 4-3 節で述べた方法で訂正し、これを B とする。A と B のうち、入力された正点字の点の並びに近い方を選び、これを最終的な訂正結果とする。

```

けいざい      名詞,一般,*,*,*,*,けいざい,ケイザイ,ケイザイ
がく          名詞,一般,*,*,*,*,がく,ガク,ガク
は            助詞,係助詞,*,*,*,*,は,ハ,ワ
EOS

```

図 8 Mecab により「けいざいがくは」を形態素解析した結果

```

えい         名詞,固有名詞,人名,名,*,*,えい,エイ,エイ
ざいがくは   名詞,固有名詞,組織,*,*,*,*
EOS

```

図 9 Mecab により「えいざいがくは」を形態素解析した結果

5 評価

4-3 節と 4-4 節の処理の効果を評価する。表 2 の「原文」の列には、スペルコレクタに入力した点字の原文を示した。ここに誤りは含まれていない。すなわち、入力をスペルコレクタを通さずにそのまま翻訳すれば原文と一致する。これをスペルコレクタにかけたときの結果と正誤を、4-1 節・4-2 節の処理のみの場合と、これらに 4-3 節・4-4 節の処理を追加した場合にわけて示している。4-1 節・4-2 節のみの処理結果を見ると、33 個のテストに対して○が 10 個である。No. 2 の「エレベータに」や No. 3 の「ケンキューシツの」のように、「に」「の」などの助詞を機械的に切り分けて訂正後につけなおすという処理でうまく訂正できている単語もあるが、「より」が切り分け対象になっていないので No. 1 では「ソバヨリ」という単語を検索して辞書にな

いため「コバヨー」という辞書にある近い単語が選ばれてしまっている。No. 31 と 32 をみると、これらはたまたま結果が正しかっただけであることがわかる。4-3 節・4-4 節の処理を追加することで、○の数は 23 まで向上している。No. 1 は「ソバ」と「より」が区別されて処理されており、No. 31 と 32 はきちんと一語として処理されている。

表 2 訂正結果とその正誤

No.	原文	4-1 節・4-2 節のみ		4-3 節・4-4 節追加	
		訂正結果	正誤	訂正結果	正誤
1	蕎麦よりうどんが好きだ	コバヨー ウドンがスキダ	×	ソバより ウドンが スキだ	○
2	エレベータに乗る	エレベータに ノル	○	エレベータに ノル	○
3	研究室の所属です	ケンキューシツの ショゾクです	○	ケンキューシツの ショゾクです	○
4	あの人が伊藤先生です	アの ヒトが イトウセンセイです	○	アノ ヒトが イトウ センセイです	○
5	こんにちは 元気ですか	コンニチわ ゲンキデスカ	○	コンニチワ ゲンキデスカ	○
6	食べながら見たので	アベナギサ シタホド	×	タベながら ミタので	○
7	危ないのに平気かしら	アバナイトに コイケアキラ	×	アブナイのに ヘイキかしら	○
8	出掛けたのかしら	イシカワケンカノー	×	デキスターかしら	×
9	食べてもいいけれど	タベレも アイコード	×	タベても イイけれど	○
10	そうとも言うけれど	セウトも イエセイド	×	セウトも イウけれど	×
11	寝てばかり	エトーヒカリ	×	エテばかり	×
12	兄へ手紙を書く	アニえ テガミを カク	○	アニえ テガミを カク	○
13	友人と一緒に	ユウジンと イッショーチ	×	ユウジンと イッショで	○
14	6時から始まるとも	ソコジカラ ツジハルトモ	×	スエジから ハジマルとも	×
15	俺はやるぜ	オレわ ヌルズ	×	オレわ ヤルぜ	○
16	負けるもんか	サクラモモカ	×	マクラモトか	×
17	寒いし暗かった	カマイシ カルカッタ	×	サムイし クラカッタ	○
18	お腹が減って死にそうだ	オナカが ヘッチ キタコーダ	×	オナカが ヘッてサイソーだ	×
19	食べられる	タグリテル	×	タべられる	○
20	知らせる	シラセル	○	シラセル	○
21	知りたがる	シリタガリ	×	シリタがる	○
22	バス乗り場へは	ビクトルバルナ	×	ビルオリバーわ	×
23	バス停へ	ガステイス	×	バステイエ	○
24	学校では	ガッコウガワ	×	ジャッコウドわ	×
25	人として尊ぶ	ヒトと シテ トートブ	○	ヒトと シて トートブ	○

26	彼にしてみれば	カレに シテ ミセバ	×	カレに シテ ミ れば	○
27	犬みたいに	イエミマイに	×	イエミマイに	×
28	良さそうです	ザソーです	×	ザソーです	×
29	アナウンサーのよう に	アナウンサーの ヨウ ニ	○	アナウンサーの ヨウに	○
30	金もなくなった	カネも ナカナダ	×	カネも ナクナッ た	○
31	おから	オから	○	オカラ	○
32	どんより	ドンより	○	ドンヨリ	○
33	秋日和	アイビより	×	アキビヨリ	○

6 まとめと今後の展望

我々が開発中のウェアラブル点字リーダーの実用性を向上させるため、圧力センサから読み取ったデータを点字の並びに再構成した際に生じる読み取りエラーを訂正するためのスペルコレクタの開発について述べた。圧力センサで点字をなぞるという動作を前提にした誤り訂正方式を検討し、点の過不足や逆さ点字に対応したスペルコレクタとこれに特化した辞書を開発した。これにより単純な語の訂正はできるようになったが、品詞への対応が不十分である。Mecab を用いて助詞・助動詞を考慮した処理を加えたことにより処理結果は大きく改善されているが、単語によってはうまく改善されない場合もあり、さらなる改善が必要である。また、現在は独立語を中心とした辞書をすべての入力に対して適用しているが、Mecabなどで品詞が適切に抽出できれば、日本語の文章として動詞がくるべき部分には動詞の辞書を、名詞がくるべき部分には名詞の辞書を、というように辞書を使い分けることが可能になる。そうすれば長い点字の文章をスペルコレクタで処理する際の精度が向上できるはずである。今後はスペルコレクタのさらなる精度向上を目指して、品詞の分析を中心に取り組んでいきたい。

表1に示したように、各国での点字の物理的な設計はほぼ同一である。このことは、圧力センサから読み取った圧力分布データを点字の1文字ごとに再構成するまでの処理はどの国でもほぼすべて同一で、点字を各国の言語に翻訳する部分だけを差し替えれば各国語対応のウェアラブル点字リーダーができることを意味している。我々は(日本式の文法で記述された)点字を日本語に翻訳するモジュールはすでに完成しており図1のシステムに組み込んでいるが、(アメリカ式の文法で記述された)点字(Unified English Braille=UEB)を英語に翻訳するモジュールも並行して開発を進めている。今後は、今回の日本語点字スペルコレクタで得られた知見を元に、UEB対応のスペルコレクタも開発を進めていく予定である。

【参考文献】

- 厚生労働省社会・援護局福祉部企画課：平成18年身体障害児・者実態調査結果，
<https://www.mhlw.go.jp/toukei/saikin/hw/shintai/06/dl/01.pdf>, 2006.
- 日本規格協会：JIS T0923:2009 高齢者・障害者配慮設計指針—点字の表示原則及び点字表示方法—消費生活製品の操作部, 2009.
- 日本規格協会：JIS T0921:2017 アクセシブルデザイン—標識、設備及び機器への点字の適用方法, 2017.
- UK Association for Accessible Formats: Standard dimensions for the UK Braille Cell ,
<https://www.ukaaf.org/wp-content/uploads/2020/03/Braille-Standard-Dimensions.pdf>, 2017.
- Australian Braille Authority: Physical Specifications for Braille ,
<https://brailleaustralia.org/about-braille/physical-specifications-for-braille/>, 2014.
- National Library Service for the Blind and Physically Handicapped of the Library of Congress: Specification 800:2014 Braille Book and Pamphlets (Section 3.2) ,
https://www.loc.gov/nls/wp-content/uploads/2019/09/Spec800.11October2014.final_.pdf, 2014.

Mecab Yet Another Part-of-Speech and Morphological Analyzer, <https://taku910.github.io/mecab/>
浅原正幸, 松本裕治: ipadic version 2.7.0 ユーザーズマニュアル, 2003.

浅原正幸, 松本裕治: NAIST Japanese Dictionary version 0.4.0 ユーザーズマニュアル, 2008.

Toshinori Sato: Neologism dictionary based on the language resources on the Web for Mecab,
<https://github.com/neologd/mecab-ipadic-neologd>, 2015.

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
感圧センサからの点字読み取りにおける補正アルゴリズム	情報処理学会第83回全国大会講演論文集 pp. 4-827--4-828, 4ZH-03	2021年3月
ウェアラブル点字リーダーの開発	第5回KMSメディカル・アーク2021 with MTO オンライン大会, S21	2021年2月