

連続音声認識の手法を応用した走査型文字入力方式*

森 大毅 (宇都宮大)

1 はじめに

ALS をはじめとする重度肢体不自由者にとって、代替コミュニケーション手段は命綱であり、同時に社会との関わりを維持し生活の質 (QoL: Quality of Life) を高めるために不可欠な道具と言える。代表的なコミュニケーション装置に、PC 上の画面から文字やフレーズを選択するものがある (Fig. 1)。可動部位が極めて限られている場合は、文字盤上でカーソルを自動走査させ、スイッチ操作によって文字などを選択する方式が利用される。しかし、カーソルの移動速度を速くすると、筋力が低下している患者は狙ったタイミングでスイッチ操作ができないため、この方法は一般に入力速度が非常に遅いと考えられている。

著者は、文部科学省科学研究費補助金特定領域研究「情報福祉の基礎」(16091104) において、走査型文字入力方式の改良の試みを行ってきた。本稿では、その成果の一部である、統計的言語モデルを利用した新しい走査型文字入力方式の基本原理について述べる。この方式は、言語モデルの利用による自動誤り訂正を行うものである。これにより、スイッチ操作に求められるタイミングの正確さが緩和され、走査速度を上げてても誤りが発生しにくくなることが期待される。

2 自動誤り訂正アルゴリズム

本手法は、言語モデルによる情報源のモデル化、スイッチ操作タイミングモデルによる通信路のモデル化、および Bayes の定理に基づく最大事後確率復号をその構成要素としており、連続音声認識の原理と多くの共通点を持っている。

Fig. 2 に統計的手法に基づく音声認識の、Fig. 3 に提案する自動誤り訂正のモデル図を示す。単語の生起確率 $P(w)$ は実際には n -gram などの文脈に依存した言語モデルによって求められる。音声認識において、音韻の音響的特徴 X の揺らぎは混合ガウス分布などによって統計的にモデル化される。走査型文字入力の枠組では、文字盤上のある位置が指定されるとき、そのスイッチ操作時刻 t が確率的な揺らぎを持つと考え、その時

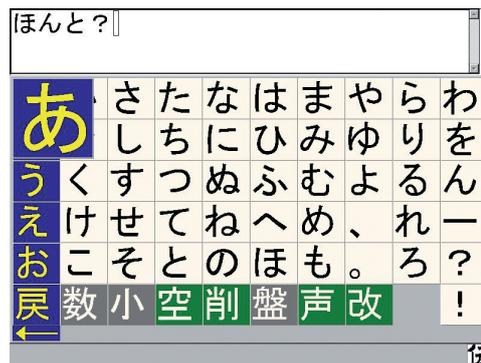


Fig. 1 「伝の心」(日立)の文字盤の例

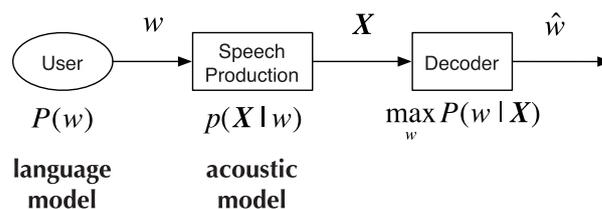


Fig. 2 音声認識のモデル

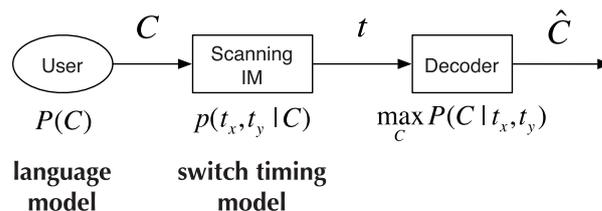


Fig. 3 走査型文字入力の自動誤り訂正のモデル

刻の分布を統計的にモデル化する。これをスイッチ操作タイミングモデルと呼ぶ。

スイッチ操作タイミングモデルは音声認識における音響モデルのような役割を果たすが、音響モデルと同じようにやはりユーザや環境に依存したモデルであることが重要である。スイッチ操作タイミングの分布は、走査速度・障害の程度や使用しているスイッチによって異なることはもちろん、健常者においても被験者による違いがかなり大きいことがわかっている。

ある文字 C が入力されるとき、列 x では時刻 t_x で、行 y では時刻 t_y でスイッチが操作される

* A novel scanning input method based on the ASR techniques. by MORI, hiroki (Utsunomiya University)

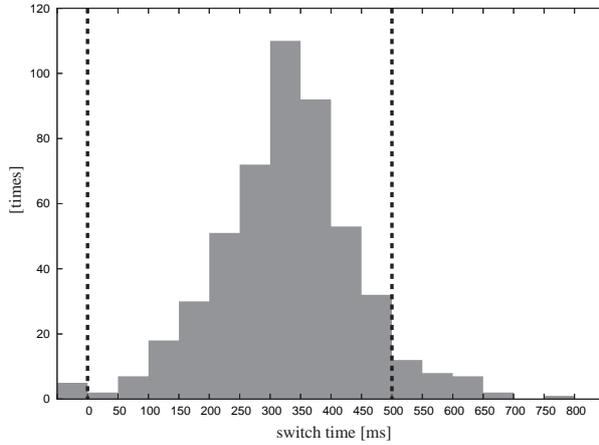


Fig. 4 ALS 患者のスイッチ操作時刻分布（ステップ間隔 500ms）

確率密度関数を $p(t_x, t_y|C)$ とする。このとき、以下の事後確率 $P(C|t_x, t_y)$ を最大にするような C の系列を出力する復号器を設計するものとする。

$$P(C|t_x, t_y) = \frac{P(C)p(t_x, t_y|C)}{\sum_{\{C, \dots\}} P(C)p(t_x, t_y|C)} \quad (1)$$

$p(t_x, t_y|C)$ はスイッチ操作タイミングモデルである。簡単のため、確率分布は C に依存せず、列方向と行方向の確率分布は独立で、それぞれ 1 次元分布 $p_x(\cdot)$, $p_y(\cdot)$ に従うと仮定すると、

$$p(t_x, t_y|C) = p_x(t_x - t_{Cx})p_y(t_y - t_{Cy}) \quad (2)$$

と近似できる。ただし、 t_{Cx} , t_{Cy} は文字 C の文字盤上の位置およびスキャン間隔によって決まる時間シフト量である。 $p_x(\cdot)$ および $p_y(\cdot)$ は何らかの確率分布で近似し、そのパラメータはスキャン法により文章を入力させる実験を行うことで推定できる。

3 スイッチ操作タイミング分布の測定

ALS 患者 1 名（57 歳男性、スキャン型意志伝達装置使用歴 10ヶ月、フットスイッチ使用）のスイッチ操作タイミングの測定を行った。入力させたテキストはイソップ童話「北風と太陽」約 500 字で、ステップ間隔は被験者にとってかなり速いと感じられる 500 ms とした。結果の度数分布図を Fig. 4 に示す。図の横軸は、入力すべき文字が属する列 / 行にカーソルが移動した瞬間を 0 ms としたスイッチ操作時刻である。図から、500 ms を超過してからスイッチが押される割合

Table 1 ALS 患者の文章入力実験結果 (%)

そのまま出力	77.7
誤り訂正あり	97.7

がかなり大きいこと、また正規性の分布形状であることがわかる。

4 文章入力実験

前出の ALS 患者 1 名に、提案する自動誤り訂正を実装した走査型文字入力システムを用いて「北風と太陽」を入力させた。スイッチ操作タイミングは単一ガウス分布でモデル化した。言語モデルは仮名 4-gram (Witten-Bell) で、毎日新聞 45ヶ月分（約 2 億字）で学習した。テストセットパープレキシティは 16.36 である。文字の探索範囲は 3×3 で、実時間動作のため仮説数一定のビームサーチを実行している。

結果を Table 1 に示す。誤り訂正をせずそのまま出力した場合の文字正解率は 77.7% であった。自動誤り訂正により、文字正解率を 97.7% まで向上できた。これより、この ALS 患者の場合には提案法が極めて有効であったと言える。

5 おわりに

走査型文字入力における入力誤りを、スイッチ操作タイミングモデルと言語モデルにより自動訂正する枠組について述べた。今回使用した言語モデルは仮名 n -gram であるが、一般の連続音声認識システムのように単語 n -gram を用いれば、さらなる精度向上や漢字仮名交じり文の生成も可能である。

ユーザ特性の違いは、使用前のエンロール行程により自動的に学習できる。ただし、実際のスイッチ操作タイミング分布は長時間使用することによる疲労のため刻々と変化する。ユーザ特性変動への動的な適応は今後の課題である。

参考文献

- [1] 森, 粕谷, 森田, 中野: “走査型文字入力におけるスイッチ操作タイミング特性の測定とその自動誤り訂正への応用,” 信学技報, WIT2006-80, 2007.