

音声入力のゲームへの応用*

7B-3

水野 正典 森谷 祐一 伏木田 勝信
NEC 情報システムズ

1 はじめに

近年、パソコンにおいてもウィンドウ環境が主流となりつつあり、以前と比べてユーザインタフェースはかなり向上している。しかしそのインタフェースは視覚的なものが主流であり、操作には依然としてキーボード、マウスが主に使用されている。

ユーザインタフェースのさらなる向上を考えると、音声によるインタフェースはかなり有効な手段となり得ることは容易に想像できる。現在も音声認識の研究は進められ、不特定話者・連続音声認識の認識率の向上が図られている。しかし、それらの研究成果はほとんど実用化されていない。[1]

その理由の一つに、研究課題が認識率や認識速度の向上といった技術的な問題に主眼が置かれ、ユーザの目的、環境に応じた応用の検討があまりなされていないことが考えられる。また、ユーザが音声入力を利用する機会がないため、ユーザの認知度が低いということも考えられる。

本稿では、音声入力をユーザにまず使ってもらうには、ゲームのように親しみのあるアプリケーションが適当であると考え、現状の技術でどの程度ゲームを楽しめるかを評価したのでその結果を報告する。

2 音声認識ソフトウェア

Microsoft¹ Windows² Sound System のボイスパイロットを音声認識ソフトウェアとして使用した。

2.1 ボイスパイロット

ボイスパイロットは Windows3.1 上で動作するアプリケーションであり、Windows の操作やアプリケー

*An application of the speech input to games, by Masanori MIZUNO, Yuichi MORIYA, Katsunobu FUSHIKIDA (NEC Informatec Systems, Ltd.)

ションのコマンドを登録しておき、キーボードやマウスの代わりに音声で操作することができる。[2]

2.2 実験環境

実験には インテル i486TM DX2 66MHz 相当の CPU を搭載するパソコンを用いた。

2.3 認識性能

認識性能については、男女混合標準パターンで 91.6%、男女別標準パターンで 93.9%、ユーザ学習後は 97.9% の認識率性能を持つとの報告がされている。[3]

2.4 応答速度

音声入力が発出され、音声入力終了、認識結果が得られるまでに必要な時間を計測した。その結果を表 1 に示す。

これによれば、ゲームに使用する単語の発声時間を 500[ms] 程度とすると、発声後その認識結果に対応する処理が実行されるまでおよそ 1[s] 経過する。これがゲームの最短応答時間となる。

表 1: 認識までの経過時間 [ms]

発声開始	発声終了	認識終了
0	269	1008
0	454	1100
0	569	1148
0	643	1523
0	1194	2119

3 ゲームの選択

実験に用いるゲームとしては時間的制約のあるものを選択することにした。これは、囲碁・将棋等の時間

的制約のないゲームに音声入力手段を提供することは容易であると考えたからである。

今回の評価では、テトリスに似たゲームを使用した。このゲームは、画面の上の方から一つずつ落ちてくる形の異なるブロックをうまく移動し、画面の左端から右端まで埋めてその行を消去し、得点を獲得というものである。

画面は縦20行から成り、ブロックは一行あたり0.7[s]の速度で落下するので、一つのブロックに最大14[s]の時間を掛けることができる。

4 評価実験

認識単語として表2の単語を辞書登録した。それぞれの単語には表に示す動作を割り当てた。各動作は表に示す複数のキー操作（'→'等は矢印キーを一回押すことを示す）に相当する。

この辞書と男女別標準パターンを用いて被験者に静かな場所でゲームを行ってもらい、実用性、操作性、娯楽性、速度について五段階で評価してもらった。

表 2: ゲーム用に登録した認識単語

単語	動作	相当するキー操作
右	右に一つ移動	→
左	左に一つ移動	←
右2	右に二つ移動	→→
左2	左に二つ移動	←←
右3	右に三つ移動	→→→
左3	左に三つ移動	←←←
右端	右端に移動	→→→→→→→→
左端	左端に移動	←←←←←←←←
右回り	右90度回転	↑
左回り	左90度回転	↑↑↑
反転	180度回転	↑↑
ダウン	落下	↓

5 評価結果

6名の被験者による評価実験では認識率94.0%であった。表3に各被験者による評価結果を示す。表3において、それぞれの項目は五段階評価であり、'1'が最も低い評価を、'5'が最も高い評価を示す。

この結果を見ると、娯楽性は比較的高く、ゲームとして楽しめていることがわかる。しかし、誤認識がゲーム進行の上で致命的な結果を招いてしまうことが

あり、全体としての評価は低いものとなった。また、応答に1[s]ほど掛かると、ゲームとしては遅く感じられるようである。

表 3: 評価結果

被験者	実用性	操作性	娯楽性	速度
A	3	2	3	3
B	2	3	4	2
C	2	2	3	3
D	1	1	4	1
E	2	2	3	2
F	3	3	4	2

6 まとめ

時間的制約のあるゲームでも現状の音声入力技術と組み合わせてある程度は楽しめる可能性のあることがわかった。しかし、誤認識や応答時間の問題など改良すべき点は非常に多い。

音声入力には、一回の発声で複数のキー入力に相当する操作が可能である。キーボード、マウス等の操作に不慣れなユーザにとって有効な入力手段である、といった利点がある。今後は、現状の問題点についてさらに調査・改良を行い、音声入力の利点を生かした、よりユーザにとって親しみやすいインタフェースを持つアプリケーションについて検討を重ねていきたい。

謝辞

日頃御指導頂く NEC 情報メディア研究所音声言語研究部の諸氏、並びに藤本正樹氏に感謝します。

参考文献

- [1] 嵯峨山：なぜ音声認識は使われないか・どうすれば使われるか？, 情処研報 94-SLP-1, Vol.94, No.40 (1994).
- [2] Microsoft Windows Sound System 1.0 ユーザーズガイド.
- [3] 畑崎他：パソコン向け音声認識ソフトウェア, 情報処理学会第 47 回全国大会論文集, 2-375 (1993).

¹Microsoft は米国マイクロソフト社の登録商標です。

²Windows は米国マイクロソフト社の商標です。