ワードプロセッサにおける音声入力実用化の条件

樺澤, 哲
松下電器産業株式会社中央研究所

高木, 英行
松下電器産業株式会社中央研究所

三宮, 真智子
鳴門教育大学学校教育学部

出版情報：The transactions of the Institute of Electronics, Information and Communication Engineers. D. J70-D (11), pp.2115-2120, 1987-11. 電子情報通信学会
バージョン：

論　文

ワードプロセッサにおける音声入力実用化の条件

正　員　樺澤　哲[†]　　正　員　高木　英行[†]　　非会員　三宮真智子[††]

Requirement for Practical Use of Voice Text-Entry for Word Processor

Satoshi KABASAWA [†], Hideyuki TAKAGI [†], *Members and* Machiko SANNOMIYA [††], *Nonmember*

あらまし　本論文では，音声入力を日本語ワードプロセッサの入力手段として実用化するための目安を主観評価実験により検討した．入力者の主観評価と強い相関をもつ物理量（性能パラメータ）を推定し，音声認識技術の開発上どの点に留意すべきかを考察した．そして，次のような結論を得た．

（1）　単音節音声入力・リアルタイム応答では認識率95％は許容できるが80％は許容できない．

（2）　また，単音節音声入力・応答時間1秒では認識率100％でも許容できない．

（3）　文節音声入力は，音節認識率が75％になれば，キー入力より好ましいと評価されそうである．

（4）　従来は，「第1位候補認識率」の向上に主眼をおいていたが，今後の開発では，それ以上に「文章作成時間」や「第n位候補認識率（認識結果の第n位までに正解が含まれる率）」や「言語処理後の認識率」にも注目すべきである．

1. まえがき

人間―機械間の望ましいコミュニケーション手段として音声認識技術が期待され，数多くの研究がなされている[(1)]．しかし，現在の技術レベルでは誤認識は避けがたく，音声入力に訂正は不可欠である．一方，キーボードによる入力（以下，キー入力）は人間―機械間の望ましいコミュニケーション手段とは言いがたく，特に不慣れな人には，多大な時間を要すると共に身体的・精神的負荷が伴う．このような，音声入力もキー入力も共に使い勝手に問題が残っている現状を踏まえて，文章作成時間[(2)]や疲労度[(3)]等の尺度で比較し，人間にとってどちらの入力方式が使いやすいかの研究もなされてきた．

しかしながら，音声認識技術開発・実用化の観点からすれば，認識方式・装置を主体とした研究だけでは不十分であると同様，被験者を主体とした研究だけでも不十分である．すなわち，認識方式・装置と人間との両者を橋渡しする，認識性能と入力者の反応との対応関係を検討することが重要である．本論文は，日本語ワードプロセッサ（以下，WP）を対象に，この対応関係を明らかにして音声認識技術開発・実用化の目安を与えることを目的にしている．

2章では単音節単位に発声する入力方法（以下，単音節音声入力）を取り上げ，認識性能を実験変数にして主観評価実験を行う．3章では文節単位に発声する入力方法（以下，文節音声入力）を取り上げ，認識率を実験変数にして主観評価実験を行う．ここでは，実用化に踏み切るためには少なくとも性能をどこまで向上させる必要があるのかを明らかにする．最後に，以上の実験を通して，入力者の評価という心理量と強い相関をもつ物理量（性能パラメータ）を推定し，音声認識技術の開発に対する留意点を考察する．

2. 単音節音声入力の検討

日本語文書の作成においては，100音節あまりを認識できれば任意の文章が作成可能である．そこで，単音節音声入力をWPの入力手段とする場合に望まれる性能，すなわち，「理想状態」と比較してそん色がない

† 松下電器産業株式会社中央研究所，守口市
Central Research Laboratories, Matsushita Electric Industrial Co., Ltd., Moriguchi-shi, 570 Japan
†† 鳴門教育大学学校教育学部，鳴門市
Faculty of School Education, Naruto University of Teacher Education, Naruto-shi, 772 Japan

性能を検討した．単音節単位と文節単位の発声では，認識率100％・リアルタイム応答とした比較実験から両者に差がないことが明らかになっている[4]．そこで，本章での理想状態は，単音節入力・認識率100％・リアルタイム応答とした．

2.1 実　　験

[実験条件] 実験条件を表1に示す．本実験で，認識率100％とした時の結果は他の認識率条件についての実験結果の基準となる．認識率95％は市販されている単音節音声認識装置のカタログ値である[1]．認識率80％は筆者らが実験的観点から現在保証できそうな値として設定したものである[6]．

応答時間については，リアルタイムと1秒を設定した．被験者は12名（男7名，女5名，22～44歳）で，WPの使用経験がほとんどない．

[実験方法] パーソナルコンピュータ（以下，パソコン）上に入力方法を模擬する入力シミュレータを実現した．パソコンにはあらかじめ単音節音声認識結果が用意されている．認識結果は一様乱数に基づいて生成した．被験者が，印刷された文章を，単音節単位に発声して，あらかじめ指定された「入力キー」を押すと，認識結果がWPのディスプレイ上に現れる．すなわち，被験者には音声で文章を入力しているように思える．被験者は，認識結果が正しければ次の音節を音声で入力するが，誤りがあれば同じ音節をもう一度発声してあらかじめ指定された「訂正キー」を押す．1回の訂正で必ず正解が得られる．本実験では，漢字変換操作は行わない．

被験者は音声認識の仕組みと入力方法・キー操作についての説明を受け，練習を経て本課題に入った．6種類（認識率3種×応答時間2種）の条件で，異なる文章を入力した．1条件での入力を終えるたびに，9項目（図1参照）についての5段階評定（図2）を行い，次の条件での入力に移った．全課題を終了した時点で，内観報告を行った．課題遂行時間は制限しなかった．

2.2 実験結果

図1に，(a)リアルタイム応答と(b)応答時間1秒に対する評定の平均値を示す．図中の横軸数字は，図2に示した5段階評定尺度の値であり，値の小さい方が評価が良い．

2.3 実験結果の検討

各評定項目ごとにt検定で有意差を調べた．結果を表2にまとめる．以下では，実験条件を[認識率，応答時間]で表す．表2(a)から[95％，リアルタイム]と理想状態との間に有意な差はほとんど認められないといえる．しかし，表2(b)から，[80％，リアルタイム]と理想状態

表1 単音節音声入力の性能評価実験の条件

要因計画：	3（認識率：100％，95％，80％） ×2（応答時間：リアルタイム，1秒）
被 験 者：	12名（男7名，女5名）
入力文章：	心理学の児童書[5]から6文章を抜き出して編集．50文節／文章（平均246.5音節／文章）
評価方法：	9項目に関する5段階評定および内観報告の分析

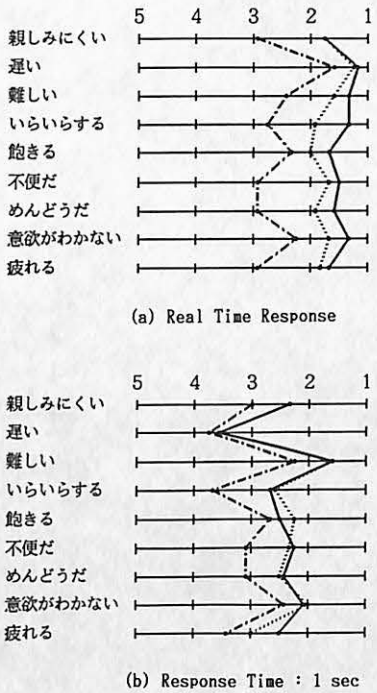

図1 単音節音声入力の評定結果
認識率100％（———）
認識率 95％（……）
認識率 80％（—·—·—）

Fig. 1 Rating result of monosyllablic voice input.
100％ accuracy（———）
95％ accuracy（……）
80％ accuracy（—·—·—）

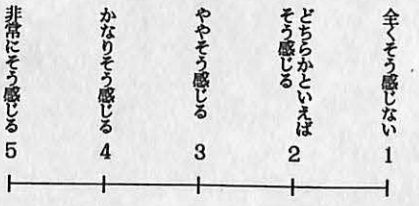

図2 5段階評定尺度
Fig. 2 Five-grade rating scale.

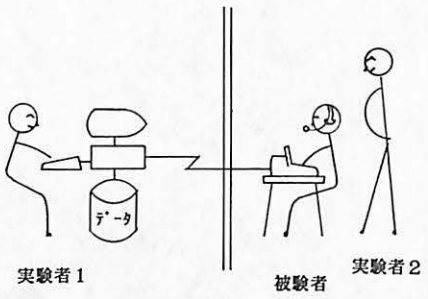

図3　実験概要図
Fig. 3　Sketch of the experiment.

表3　キー入力と文節音声入力の評価実験の条件

要因計画：	2(入力方法：かなキー，文節音声) ×3(音節認識率：65％，70％，75％)
被験者：	各音節認識率ごとに9名(男2名，女7名)ずつ計27名
入力文章：	心理学の児童書[5]から抽出・編集した1文章(55文節・239音節)
評価方法：	9項目に対する5段階の評定および内観報告の分析

3.3　実験結果

認識率条件ごとに，キー入力・文節音声入力の各々に対する評定の平均値を図4に示す．

3.4　実験結果の検討

t検定を用い，各評定項目ごとにキー入力と音声入力との有意差を調べた．結果を図4の右端に示す．本評価実験結果から音節認識率が75％になれば，音声入力の方がキー入力より好ましいと評価されていると判断できそうである．

三つの認識率条件におけるキー入力と文節音声入力との比較において，入力文章・実験順序等の条件が一定なので，キー入力の評定結果は一定であることが期待される．しかし，認識率75％の評定結果では，他の認識率条件の場合に比べてキー入力の評価が悪くなっている．この点について，認識率条件に対応した各被験者グループ間でキー入力に能力差があったかどうかを，実験での文章作成時間で考察する．文章作成時間が正規分布に従うと仮定すると，キー入力での文章作成時間は図5(a)に示される．図から被験者グループ間に差があるとは思われない．よって，認識率の向上に伴って文節音声入力の評価が良くなったことが相対的にキー入力の評価を下げた，と考えられる．その他に実験時期・時間帯によって被験者間に疲労感等の差が若干生じたことも考えられるが，この影響はキー入力・文節音声入力の両方に及んでいるはずである．以上の考察から，キー入力と文節音声入力との評定差は信頼できる情報と考えられる．

また，各認識率条件間でキー入力と文節音声入力とを比較してみると，図5(b)の平均文章作成時間の差は，図4の評定差に対応しているように見える．

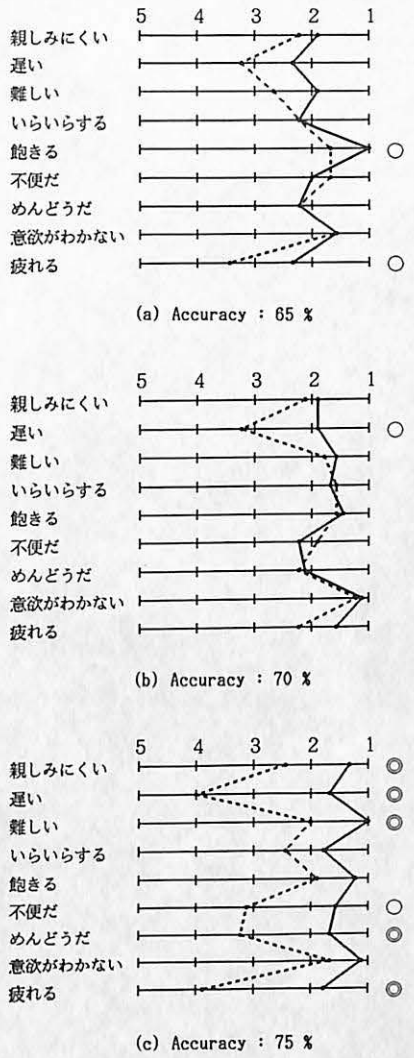

図4　キー入力(……)と文節音声入力(———)との評定結果
右端の記号はt検定による有意差
◎：1％の危険率で有意差あり
○：5％の危険率で有意差あり

Fig. 4　Rating results of keyboad input (……) and BUNSETSU†-voice input (———).
The result of t-test
◎ : The difference is significant ($p<0.01$)
○ : The difference is significant ($p<0.05$)
(† BUNSETU is a portion of Japanese sentence between pauses which are generated by natural breathing when the Japanese reads aloud a Japanese sentence.)

表2　t検定結果

	(a)	(b)	(c)	(d)
親しみにくい		○		
遅　い			◎	◎
難しい		◎		○
いらいらする		◎	○	
飽きる		○	○	
不便だ		◎	○	
めんどうだ		◎	○	
意欲がわかない		○	○	
疲れる		◎	○	

(a)　認識率：100％ vs. 95％，リアルタイム応答
(b)　認識率：100％ vs. 80％，リアルタイム応答
(c)　認識率：100％，応答：リアルタイム vs. 1秒
(d)　認識率：100％•1秒 vs. 認識率：80％•リアルタイム
◎：危険率1％で有意差あり
○：危険率5％で有意差あり

との間には有意な差が認められる．表2(c)から，[100％，1秒]と理想状態との間にも有意な差がある．また，表2(d)において，一部の評価項目で有意差が見られるものの全体的にみれば，[100％，1秒]と[80％，リアルタイム]とがほぼ同じ評価であると考えられる．すなわち，リアルタイム応答では認識率95％は許容できるが認識率80％は許容できない，応答時間1秒では認識率100％でも許容できない，といえる．

3.　音声入力の実用化に必要な性能

前章での実験結果を得たのと同じ時期に，単音節音声入力に否定的な報告がなされ[3]，筆者らが前章で得た結果と一致する部分があった．また，文章入力を実現するには少なくとも文節あるいは句単位の発声が必要である[7]，との意見もある．

そこで，本章では対象を文節音声入力とし，実用化に必要な認識率を明らかにする．筆者らは，実用化の条件とは「文節音声入力がキー入力より好ましいと評価されること」であると考え，実験を行った[8]．

3.1　認識シミュレータ

2.では，一様乱数に基づいて認識結果を作成したが，本章では「認識シミュレータ」を作成して，より現実に近い認識結果で実験した．

この認識シミュレータは，認識部と言語処理部からなる．認識部は，筆者らが試作した単音節音声認識装置から得た認識結果と同じ誤り傾向をもつ．単音節音声認識装置内部の標準パターンと入力単音節のパターンとの距離が正規分布に従って分布するという仮定にたって[9],[10]，分散を変化させることで認識率を制御した．言語処理部は，認識部から得られた単音節ラチスに基づいて，日本語として存在し得る文節のみを出力する[11],[12]．

なお，本章での音節認識率とは，認識シミュレータの認識部で得られる第1位候補認識率を単音節あたりに換算したものである．被験者に提示する言語処理後の認識率は音節認識率65％，70％，75％に対応して，それぞれ80.3％，82.8％，89.1％である．

3.2　実　　験

[実験条件]　実験条件を表3に示す．予備実験では，60％程度の音節認識率の文節音声入力はキー入力に若干劣る†との結果を得た[13]．そこで，音節認識率を65％，70％，75％に設定し，キー入力と文節音声入力とを比較評価した．被験者は，各音節認識率ごとに9名(男2名，女7名，18～20歳)で，WP経験が浅く「かな入力」には慣れていない．

[実験方法]　実験装置は液晶7行表示の逐次変換方式WPであって，パソコンに接続されている．キー入力と文節音声入力を同じWPで評価した．キー入力は「かな入力」で行う．文節音声入力は文節単位に連続音声で入力する．被験者に提示された認識結果が正しくない場合は，訂正キーで順次次候補を選択して正しい文節を得る．正解が候補にない場合は，再発声するか「かな入力」する．しかし，再発声しても，実際にはあらかじめ用意されたシミュレータの出力を順次繰り返して提示するだけであり，いずれは「かな入力」しなければならない．

実験の様子を図3に示す．文節音声入力の場合には，実験者1が被験者の発声に合わせて，パソコンにあらかじめ用意されている認識結果を提示すると共に，WPに接続されているパソコンを通じてキー操作を監視記録する．実験者2は，被験者のキー操作・発声を監視指導する．被験者には，シミュレータの存在やWPがパソコンに制御されていることを知らせず，あたかも音声で文章が作成されているかのように思わせる．

実験は，キー入力→文節音声入力→両入力に対する評定→実験全体を通しての内観報告，の順に進めた．評定では，9項目(図4参照)に対して5段階評定(図2)を行った．

† 但し，文節音声入力がキー入力に比べ不利な実験条件であった．

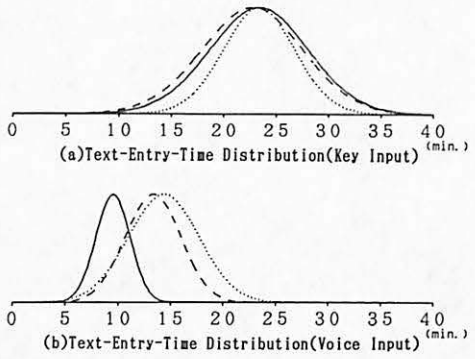

図5　被験者グループ別文章作成時間分布
Fig. 5　Text-entry-time distribution of each subject group.
75 % accuracy group (———)
70 % accuracy group (------)
65 % accuracy group (·····)

更に，三つの認識率条件とも，文章作成時間の平均値および分散は，文節音声入力の方がキー入力より小さい．このことから，文節音声入力の方が，「誰もが，速く」入力でき，「取っ付きやすい」と考えられ，一般に指摘されている音声入力の特長に一致する．

4. 今後の開発への指針

図4において，実験条件の認識率と心理的な評価の変化とは必ずしも一対一対応しているとはいえない．もし，この心理量に一対一対応する物理量が見つかれば，今後の音声認識技術の開発においてどの点を第1に改良すべきかが明らかになるであろう．本章ではこの考察を行う．

総合的な評価の変化量は，図4における二つの入力方法の，各評価項目ごとの差を各々に重み付けすることで表せる．この重み付けは，主成分分析で成分得点(因子得点)を求めることで行った．第1因子の寄与率が78.8％であったため，総合的な評価を第1因子に対する成分得点で代表させると，心理量の変化は表4のようになった．認識率条件(65％→70％→75％)に伴って変化する物理量の幾つかについて，物理量変化と心理量変化(表4)との相互相関係数を表5に示す．但し，第 n 位候補認識率とは，第 n 位認識候補までに正解が含まれる率を意味している．

比較する実験条件がわずか3点であるので，結論づけるのは今後の研究を待たねばならないが，表5から次のようなことがいえる．すなわち，実験変数とした「第1位候補認識率」よりも，「文章作成時間」，「第 n 位候補認識率」および「言語処理後の認識率」の方が，評価という心理量に対応している．このことから，従来，音声認識技術開発では第1位候補認識率の向上に主眼をおいて評価していたが，それ以上に，表5で示したような視点に着目すべきである，といえる．例えば，第1位候補認識率が低くても第2位候補認識率が高ければ，その逆の場合よりユーザーの評価は高くなることもあり得る．

表4　キー入力と文節音声入力の評価差の心理量

認識率65％	認識率70％	認識率75％
1.5	1.3	4.3

表5　物理量と心理量の相互相関係数

物　理　量	相互相関係数
第1位候補認識率	0.83
第2位候補認識率	0.95
第3位候補認識率	0.90
第4位候補認識率	0.98
第5位候補認識率	0.93
言語処理後の認識率	0.94
誤認識による再発声回数	0.90
訂正キーで正解を探したが存在せず再発声した回数	0.79
文章作成時間	0.99

5. むすび

本論文では，音声認識技術開発・実用化の目安を得ることを目的とし，WPを対象に，音声認識性能と入力者の反応との対応関係を検討して，次のような結論を得た．

(1)　単音節音声入力・リアルタイム応答では認識率95％は許容できるが80％は許容できない．

(2)　また，単音節音声・応答時間1秒では認識率が100％でも許容できない．

(3)　文節音声入力は，音節認識率が75％になれば，キー入力より好ましいと評価されそうである．

(4)　従来は，「第1位候補認識率」の向上に主眼をおいていたが，今後の開発では，それ以上に「文章作成時間」や「第 n 位候補認識率」や「言語処理後の認識率」にも注目すべきである．

謝辞　本研究を進めるにあたり，徳島県立城南高等学校教諭吉谷篤志ならびに鳴門教育大学院生三尾忠男両氏に多大な御助力御助言を賜った．ここに記して深謝する．また，本研究の機会を与えて頂いた関係各位にも感謝する．

文　献

(1) "特集：音声入力装置の商品化競争―いよいよ来る音声入出力時代―", ビジネス・コミュニケーション, **21**, 9, pp 36-56 (昭 59-09).

(2) J. D. Gould, J. Conti and T. Hovanyecz : "Composing letters with a simulated listening typewriter", Communication of the ACM, **26**, 4, pp. 295-308 (1983-04).

(3) 北原, 中山, 遠藤："音声による文章入力に関する人間工学的検討", 第 30 回情報処理全大前期, 6N-2 (昭 60-03).

(4) M. Sannomiya, S. Kabasawa, H. Takagi and A. Yoshiya : "A study on voice recognition as human interface for Japanese word-processor", 2nd Symposium on Human Interface 1421, pp. 545-552 (Oct. 1986).

(5) 加藤　秀著, 拓殖秀臣監修："少年少女　脳と心の科学", 童心社 (昭 58).

(6) 樺澤, 楠原, 松井, 相良, 前原, "単音節認識における誤認識要因の検討", 音講論, 3-1-2 (昭 58-10).

(7) 千葉成美 ："音声情報処理装置の市場動向", 音響誌, **42**, 12, pp. 959-964 (昭 61-12).

(8) 三宮, 高木, 樺澤, 三尾："主観評価実験による音声入力とキー入力との比較", 音講論, 3-5-10 (昭 62-03).

(9) 阿部, 秦野, 福村："辞書を利用する文字認識系の能力の評価", 信学論(C), **52-C**, 6, pp. 305-318 (昭 44-06).

(10) 外川, 田中, 上田, 金原："模擬単音節認識による辞書照合の検討", 音講論, 1-4-2 (昭 57-03).

(11) 高木, 中嶋, 楠原, 前原："日本語統計情報の音声認識への応用", 音講論, 1-4-21 (昭 60-03).

(12) 高木, 楠原, 坪香："音声日本語文入力における日本語統計情報の評価", 音講論, 1-1-25 (昭 61-03).

(13) 高木英行："音声ワードプロセッサの主観評価", 第 2 回ヒューマン・インタフェース・シンポジウム　2242, pp. 357-360 (昭 61-11).

(昭和 62 年 3 月 30 日受付, 6 月 1 日再受付)

樺澤　哲

昭 50 阪大・工・通信卒. 昭 55 同大学院博士課程了. 工博. 同年松下電器産業㈱入社. 現在, 同社中央研究所に勤務. 音声信号処理・自然言語処理の研究開発に従事. 情報処理学会・日本音響学会各会員.

高木　英行

昭 54 九州芸工大・音響設計卒. 昭 56 同大学院修士課程了. 同年松下電器産業㈱入社. 現在, 同社中央研究所に勤務. 音声処理の研究に対し, 言語処理・主観評価・認識の面からアプローチ, 日本音響学会会員.

三宮真智子

昭 53 阪大・人間科学部卒. 昭 58 同大学院博士課程了. 昭 59 鳴門教育大・学校教育学部助手, 現在に至る. 学術博. 人間の情報処理メカニズムの研究に従事. 日本心理学会, 日本教育心理学会, 日本基礎心理学会, 日本教育工学会各会員.