音声言語情報処理 14－9
(1996. 12. 12)

データ入力システムの性能と使用感に関する調査

山本 寛樹　小坂 哲夫　山田 雅章　小森 康弘　藤田 稔

キヤノン（株）情報メディア研究所

〒211 川崎市幸区鹿島田890-12

E-mail: hiroki@cis.canon.co.jp

あらまし　一般に実環境で音声認識を用いるとデータベースを用いた実験に比べ，性能が低下することが知られている．しかし，その要因については現段階で全てが明確になっているわけではない．また，実際のサービスにおいて，音声認識システムの性能とユーザの使用感の相関はわかっていない点が多い．これらを調査するため，我々は，データ入力システムを試作し，被験者14名によるフィールドテストおよびアンケート調査を行なった．本稿では，試作したシステムの性能評価およびシステムの性能とユーザの使用感の相関を調査した結果を報告する．

キーワード　音声認識，データ入力システム，フィールドテスト，使用感

Performance of data input system and users' impressions of the system.

Hiroki YAMAMOTO, Tetsuo KOSAKA, Masayuki YAMADA, Yasuhiro KOMORI and Minoru FUJITA

Media Technology Laboratory, Canon Inc.

890-12 Kashimada, Saiwai-ku, Kawasaki-shi, KANAGA 211, Japan

E-mail: hiroki@cis.canon.co.jp

Abstract　Generally, performance of speech recognition evaluated in the real world becomes worse than that evaluated on the databese. But not all factors of the performance degradation are cleared. Additionally, we don't know well what factor of speech recognition technology promotes comfortable use of the system. In order to investigate these factors, we made a data input system with speech input and performed a field-test with a questionnaire survey on 14 subjects. This paper describes about the results of the field-test and the correlation between the performance of the speech recognition system and the users' impressions of the system.

key words　speech recognition, data input system, field-test, users' impressions

1 はじめに

従来我々は音声認識をデータベース上のデータで評価してきた．しかし，一般に実環境で音声認識を用いると性能が低下することが知られている．したがって，音声認識システムを開発するにあたり，実環境において生じる問題点を整理しなければならない．また，実際のサービスにおいて音声入力が他の入力手段に比べ，どのような場面で優位性を示すか，音声認識システムのどの性能がユーザの使用感に影響するか，などわかっていない点も多い．

フィールドテストを行なってこのような問題を調査した結果が各研究機関から報告されている [1, 2, 3]．武田は実環境における音声認識性能低下の問題点を指摘し，改善策を提言している [1]．森島は音声認識のユーザの使用感や他の入力手段 (プッシュホン) との使用感の比較などのアンケート結果を報告している [2]．吉岡は正解候補を複数出力する音声入力の優位性をタスク達成時間の観点から報告している [3]．しかし，どのような技術要素が重要であるか，現段階ではすべてが明確になっているわけではない．また，ユーザの使用感と音声入力システムの性能の相関は報告されていない．

我々は，既存のエンジンを実環境で用いた場合の性能およびシステムの性能とユーザの使用感の相関を調べるため，データ入力システムを試作し，被験者 14 名によるフィールドテストおよびアンケート調査を行なった．試作したシステムは，社内の伝票を音声入力で作成するシステムで，認識部には我々が開発を進めてきた不特定話者・連続入力の音声認識エンジンを用いている．認識対象語は総数で 2,557 語である．その他の特徴として，認識結果の複数候補出力の利用，動的文法切替え，不要語の棄却などがある．

本稿では，フィールドテストの結果およびアンケート調査の結果をもとに，本システムの性能を評価し，ユーザの使用感と音声入力システムの性能の関わりについて述べる．

以下では，まず2章で試作したデータ入力システムについて述べる．次に3章でフィールドテストの概要を述べる．4章ではフィールドテストで得られた結果から本システムの性能の評価を行ない，現状の問題点および改善すべき点を述べる．5章で被験者に行なったアンケート調査を基に，音声認識システムの性能とユーザの使用感の関わりについて述べる．

2 データ入力システムの概要

本システムは，社内で用いている交通費の申請用紙をマウスと音声入力を用いて作成するデータ入力システムである．

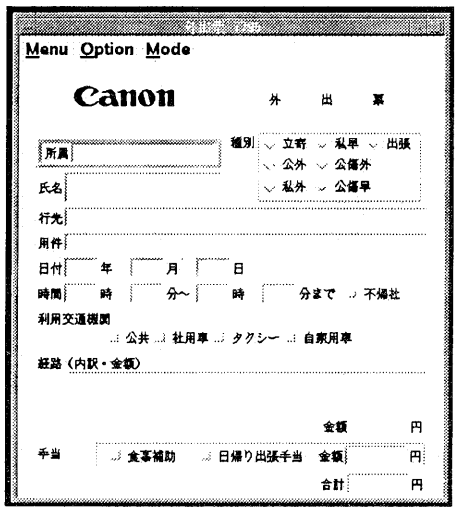

図 1: データ入力システム

2.1 システムの構成

本システムは計算機・ディスプレイ・キーボード・マウス・マイクで構成される．計算機は HP 9000 J210 を使用し，マイクはワイヤレスのハンドマイクを用いた．

2.2 入力方法

本システムでは，入力開始と同時に申請用紙を摸式した図 1のようなウィンドウを表示する．用紙作成に必要な入力項目は，所属・氏名・外出の種別・行先・用件・日付・時間・利用交通機関・経路・手当の 10 項目である．全ての項目について音声による入力が可能である．また，他の入力手段として，外出の種別，利用交通機関はマウスによるボタンクリック，その他の項目はキーボードによる入力が可能である．入力は所属から順番に各項目ごとに行なう．音声入力を用いる場合は，メインウィンドウの他にレベルメータや認識結果の複数候補を表示するウィンドウが表示される．多くの項目は単語入力であるが，日付・時間・経路の三項目では連続入力が可能である．認識結果は図 1のウィンドウの当該欄に表示し，尤度の高い順に上位 5 候補を別のウィンドウに表示する．表示された第二位以下の候補の選択にはマウスを用いる．項目の移動は，キーボード入力で改行するか，マウスで移動先の項目をクリックするか，音声による指示で行なう．また，誤認識によりユーザが意図しない動作をシステムが行なった場合に，認識前の状態に戻す訂正機能を設けた．本システムでは，データ入力以外にシステムの操作を指示するコマンドを音声入力できる．『次』『戻る』や,『所属』『氏名』などの項目名を発声することにより項目の移動ができる．他に音声入力できるコマンドは『訂正』や入力を終了する『終了』などがある．音声による訂正指示や入力終了の指示に対しては，『はい』『いいえ』でユーザの確認をとる．

2.3 音声認識部

音声の取り込みは常時行ない，一般的に用いられている短時間パワーと零交差回数を用いて音声区間を検出している．また，音声区間の始端検出と同時に音響分析・音声認識を駆動し，ほぼ実時間での処理を行なっている．音声の終端を検出した時点から認識結果が表示されるまでの間は音声取り込みを中断している．

音響分析はサンプリング周波数16kHz，フレーム周期10msec，窓幅25.6msec，プリエンファシス0.97，特徴量としてLPCメルケプストラム12次+ΔLPCメルケプストラム12次+Δ対数パワーを用いている．

音声認識に用いたHMMは，3状態6混合・対角化共分散行列の混合連続出力分布型で，右環境依存型(243種類)である．HMMの学習には，音響学会(話者64名，各150文)とATR(話者40名，各216単語+50文+15数字)データベースを用いた．

音声認識システムの文法は，BNF表記文法を有限状態オートマトンに展開している．認識対象語数は2,557で各入力項目ごとに文法を替えている．それぞれの文法の認識語数は表4に示す．

探索アルゴリズムはtree-trellis based searchによりN-Bestを求める方法を用いている．前方向の探索では，各時点における最大尤度より一定値を下まわる閾値により枝を刈るBeam Searchを併用している．

また，筆者らが提案した高速尤度計算方法IDMM+SQ[4]と未知語処理[5]を用いた棄却機能を組み込んでいる．IDMM+SQで用いるスカラー量子化コードブックのサイズは16である．発声を棄却する場合は，メッセージをウィンドウ上に表示してユーザに棄却したことを知らせる．

3 フィールドテストの概要

3.1 テスト方法

各被験者には，まずはじめにデータ入力システムの使用方法を15分で説明した．続いて，テストで用いるタスクセットを用いて音声入力の練習を20分～30分間行なった．フィールドテストでは被験者に対し，表1に示す5つの課題を与えた．課題1～3では主として音声入力，課題4ではキーボード入力，課題5では手書きによる用紙作成を課題とした．音声入力を行なう課題では，繰り返し発声しても認識されない場合はそのデータの入力を放棄しても良いことにした．表1中の'放棄'は最低限試行してもらう音声入力の回数である．また，課題4のキーボード入力の実験を行なうにあたり，事前に入力対象となる単語をかな漢字変換辞書に登録した．全ての課題が終了した時点で，被験者にアンケートを行なった．

表1: 実験

	入力	N-Bestの利用	マウスによる項目の移動	放棄
課題1	音声	○	×	3回
課題2	音声	3回目以降○	×	5回
課題3	音声＋マウス	○	○	3回
課題4	キー＋マウス	—	○	—
課題5	手書き	—	—	—

3.2 被験者

被験者は筆者らを含め，合計14名である．内訳は，音声研究者4名(R1～R4，全て男性)，男性5名(M1～M5)，女性5名(F1～F5)である．被験者は技量に個人差はあるが全員キーボード入力の経験がある．また音声認識システムの利用経験は，『何度も使ったことがある』4名(R1～R4)，『過去に数度』6名(M1～M5,F1)，『今回が初めて』4名(F2～F5)である．

3.3 タスク

テストでは，予め作成した申請用紙を5種類のタスク(t1～t5)を用意し，表2に示すタスクセットを作成した．表には各タスクセットを用いた被験者をあわせて示す．

表2: タスクセット

タスクセット	タスク					被験者
	t1	t2	t3	t4	t5	
O-set	○	○	○	○	○	R1～R4
A-set	○	○	○	—	—	M1,F1
B-set	○	—	—	○	○	M2,F2
C-set	—	○	○	○	—	M3,F3
D-set	○	○	—	—	○	M4,F4
E-set	—	—	○	○	○	M5,F5

3.4 収録したデータ

実験では入力開始から終了までマイクから集音された全ての音を収録し，これとは別に切り出された音声区間の波形を収録した．また，入力に費やした時間，音声認識結果，マウスイベントの履歴などを時間情報を添えて記録した．

4 フィールドテストの結果

4.1 区間検出

被験者14名の発声内容の内訳を表3に示す．

表3では，正しく検出されなかった発声を『検出洩れ』と『検出区間誤り』とに分類した．『検出洩れ』は被験者の発声が検出されなかった回数を示している．検出洩

表 3: 発声内容内訳

総発声数	5470	
検出洩れ	53	(0.97%)
検出区間誤り	72	(1.32%)
正確に検出	5345	(97.71%)
データの脱落	89	(1.63%)
被験者発声ミス	124	(2.27%)
正しい発声	5132	(93.82%)
誤検出	8	

れになった 53 発声の内訳は，項目の移動を音声で行なう時に用いる『次』が 37，入力終了時にユーザの確認をはい/いいえで問う時の『はい』が 10 でほとんどを占めた．音節数が少ないため発声継続時間が短いことや，語頭母音の脱落などが検出洩れの原因と考えられる．『検出区間誤り』は検出した音声区間に誤りがあった回数である．全て発声の始端検出を失敗して語頭が切れたデータだった．検出区間誤りした発声は『タクシー』(13),『アクセス』(9),『不帰社 (フキシャ)』(7),『行先』(7) などで，誤りの原因は語頭母音の無声化や脱落により短時間パワーが検出のための閾値を越えなかったためと推測される．

正しく検出したデータの中には，被験者の発声ミスや音声を取り込む段階で音声データの一部が脱落した数を含んでいる．被験者が文法外の言葉を発声した場合，言い淀んだ場合,『あっ』『えー』など不要語をともなって発声した場合を発声ミスとして数えた．データの脱落の原因の一つに，音声区間の終端検出から認識結果を表示するまでの間に被験者が発声した場合があった．

雑音を音声と誤って検出した場合を『誤検出』とした．主な原因は，マイクのスイッチの on/off で生じた雑音であった．

被験者の発声を正しく検出する割合は，97.71%で，音声区間検出の段階で約 2% の発声に影響が出ている．データの脱落まで含めると，システム側の問題で入力時に約 4%のデータに欠損が生じている．

4.2 音声認識率

本実験の音声入力では，連続発声，単語発声が混在しているので，音声認識率は以下の式で計算した．被験者の発声ミスや区間検出誤りなど不完全な入力を棄却した場合は正解として数えた．

$$\text{音声認識率 (\%)} = \frac{\text{正解数}}{\text{音声区間検出数}} \times 100$$

全課題を通じての平均認識率は，82.69%である．不完全なデータを取り除いた正しい入力のみでの認識率は 84.02%，上位 5 位内の認識率は 90.86%であった．

入力項目ごとの認識率を表 4に示す．表では，データを入力するために発声した時の認識率とコマンドを発声した時の認識率をあわせて示す．

表 4: 入力項目ごとの認識率 (%)

入力項目	認識語彙数	認識率 (%) 総合	データ	コマンド
所属	104	91.36	89.71	99.16
氏名	241	92.31	87.63	99.24
外出の種別	41	94.78	93.08	96.64
行先	146	93.97	90.23	98.58
用件	61	98.23	96.95	100.00
日付	123	77.78	67.20	100.00
時間	113	68.87	61.43	96.85
利用交通機関	32	92.28	87.20	99.17
経路	1630	74.75	69.45	96.90
手当	59	95.09	93.44	96.30
はい/いいえ	7	99.00	——	99.00
合計/平均	2557	82.69	74.60	98.07

項目別の認識率では，数値の入力が伴う日付，時間，経路などの認識率が低い．また，システムの操作を指示するコマンドの認識率が高かった．数値の入力でよく見られた誤認識は,『280 円』→『180 円』,『7 月 (シチガツ)』→『1 月』,『14 時 50 分まで』→『10 時 50 分まで』などである．本システムでは，単語境界で音素モデルを用いており，音素環境依存モデルを用いていない．このため，数値の認識性能が低くなったと考えられる．

次に課題別の認識率を図 2に示す．

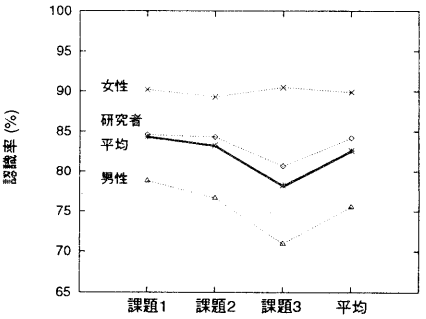

図 2: 課題別の音声認識率

課題別では，課題 3 の時，男性被験者および研究者の被験者の認識率が低下している．この原因は，課題 3 ではマウスによるシステムの操作回数が増え，認識率の高いコマンドの発声回数が減少したためと考えられる．本システムの音声認識部は，データベースを用いた実験では，男女性で認識率の差はあまりなかった [4] が，フィールドテストでは全課題を通じて女性の認識率が一番高かった．これは，研究者や男性が日常の会話のように発声していたのに対し，女性の発声が丁寧であったためと考えられる．発声の丁寧さの一つの指標として発声速度を調べ，認識率との相関を図 3に示す．

図 3から発声速度は，女性<研究者<男性の順に早くなっている．また，発声速度の早い被験者は認識率が下

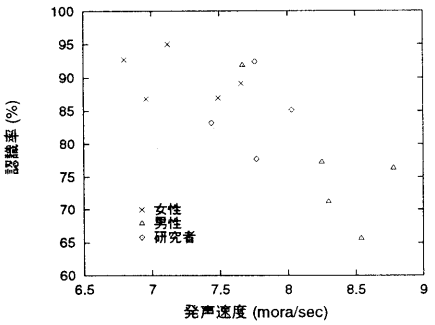

図 3: 発声速度と音声認識率の関係

がる傾向がある.

この結果から，現状では，丁寧な (適当な発声速度の) 発声で入力した方が認識率が良いことがわかる.

4.3 アクセプト率

被験者の発声に対するシステムの挙動を調べるため，アクセプト率を以下の式で計算した.

$$\text{アクセプト率}(\%) = \frac{\text{発声をアクセプトした数}}{\text{発声数}} \times 100$$

アクセプト率の計算では，被験者の発声ミスやシステム側のミスであっても，発声通りシステムが動作しない場合はアクセプト数に数えなかった. 全課題・全被験者の平均アクセプト率は 79.91%であった.

データを入力する際に，誤認識した場合は被験者は同一内容を繰り返し発声する. この同一発声内容を繰り返した場合のアクセプト率の推移を図 4 に示す. 横軸は，同一内容の発声の繰り返し回数で，例えば『2nd』は一回目に発声して間違った内容を 2 回目に発声した時のアクセプト率を示す.

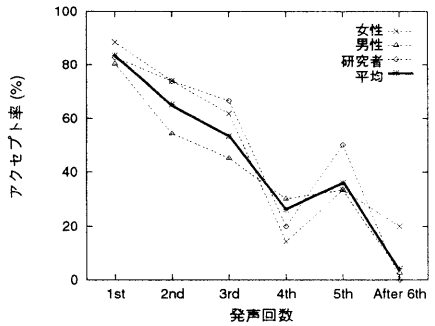

図 4: 同一発声を繰り返した時のアクセプト率の推移

図 4から，繰り返し回数を重ねるごとにアクセプト率が低下していく傾向がわかる. 繰り返し発声する場合は認識しにくい内容であることが多く，3 回目の発声でアクセプト率は 50%近くまで低下する.『After 6th』は 6 回目以降の発声の合計であり，アクセプト率は 4.17%(48 発声中アクセプト 2) である.

これらの結果から，2 度 3 度繰り返して入力してもアクセプトされない場合は，何度繰り返し入力してもアクセプトされる可能性は少ないことがわかる. なかなか認識されない語を減らすには音声認識性能の向上が必要であるが，あわせてシステム側で何らかの方法で同一発声が繰り返されていることを検知し，一定回数以降は別の入力方法に切替えるなどの方策が必要と考える.

4.4 棄却処理の性能

総発声数 5,470 のうち，検出されなかった発声数を引き，誤検出数を加えた 5,425 の音声が認識部に入力された. 5,425 のデータのうち，システム側の問題で不完全なデータが 169，被験者の発声ミスなどが 124，合計 293(5.4%) の不完全な音声が認識部に入力されている. 不完全データの入力に対する認識部の振るまいを表 5に示す. 括弧内の数字はそれぞれの項目における割合 (%) である.

表 5: 不完全データに対するシステムの振舞い

	被験者のミス		システムのミス		合計	
総数	124	(100.0)	169	(100.0)	293	(100.0)
棄却	81	(65.32)	30	(17.75)	111	(37.88)
認識正解	1	(0.01)	66	(39.05)	67	(22.87)
誤認識	42	(33.87)	73	(43.20)	115	(39.24)

システム側のミスは，40%をそのまま認識して正解しているが，棄却できるのは 18%に留まる. 特に，入力時に誤検出した 8 個の入力は全て誤認識した. 逆に，被験者のミスはその 65%を棄却している. 被験者の発声ミスの内容と棄却した数を表 6に示す.

表 6: 被験者の発声ミスの内訳と棄却率

ミスの内容	発声数	棄却数	棄却率
言い淀み	48	34	70.8
入力項目の間違い	64	43	67.2
不要語	10	2	20.0
読み間違い	2	2	100.0
合計	124	81	65.32

棄却処理により，正しく入力されたデータを棄却する場合もある. 表 7に棄却されたデータの内訳および棄却処理機能を外した時の認識結果を示す.

表 7の右側は，棄却された 317 発声の棄却処理を行なわない場合の認識結果である. 括弧内はビームで刈られて認識結果が出力されなかった (棄却に等しい) 数である. この結果から，棄却処理により『なし』の場合で正しく認識されていた 96(=108-12) 発声を間違えて棄却していることがわかる. 逆に,『なし』の場合に誤認識していた 96(=68+28) の不完全データを棄却していることもわかる.

表 7: 棄却されたデータ内訳

棄却処理	あり		なし	
データの内訳	正解	誤認識	正解	誤認識
被験者の発声ミス	81		13(11)	68
システムのミス	30		2(1)	28
正しい入力		206	93	113
合計	111	206	108 (12)	209

以上の結果から，我々のシステムでは，棄却処理を入れても入れなくてもシステム全体の認識率に大きな影響は出ないと考える．今後，さらに棄却処理の性能を向上する必要がある．同時に，現状のまま棄却処理を用いるには，致命的な誤認識を減らす利点と尤度が小さいながらも正解している認識結果を棄却する欠点を考慮し，入力項目による使い分けなどを考えなければならない．例えば，コマンドを受け付けるような項目では棄却処理を用いて誤認識による誤操作を防ぎ，データ入力のみを受け付ける項目では棄却処理を用いない等の使いわけが考えられる．

4.5 タスク達成率

タスク達成率は，以下の式で計算した．

$$\text{タスク達成率}\,(\%) = \frac{\text{正しく入力できたデータ数}}{\text{入力すべきデータ数}} \times 100$$

発声回数ごとのタスク達成率を図5に示す．横軸の数字は発声回数を示し，『数字+N』は N-Best を利用した場合を示す．例えば『2+N』は 2 回目の発声の認識結果の N-Best を利用した場合である (N-Best 候補に正解が含まれても被験者が利用しなかった場合は数えていない)．『Last』は最終的なタスク達成率である．誤認識時の，再発声回数の上限に制約を設けなかったため (下限は設定した)，ほとんどの被験者が入力できるまで発声を繰り返していた．

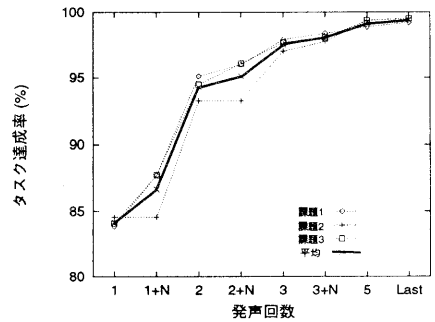

図 5: タスク達成率

最終的なタスク達成率は，全課題平均で 99.41%であった．達成されなかった部分は，音声入力で何度発声しても認識できずに放棄したデータや，誤認識したまま被験者が気づかずに入力を終了したデータである．

全課題平均では，一回目の発声で 84.3%のデータが正しく入力されている．達成率は『3+N』の時点で，98.1%になり，5 回目の発声終了時で 99.2%に達する．また，『1』と『1+N』の差,『1+N』と『2』の差を比べると，N-Best の利用よりも発声し直しす方がタスク達成率の伸び率が高いことがわかる．

以上の結果から，98.1%のデータが 3 回目の発声までに入力されたことがわかった．また，N-Best はタスク達成率に寄与するが，発声し直した方が入力され易いと言える．

4.6 タスク達成時間

図 6に課題別のタスク達成時間を示す．

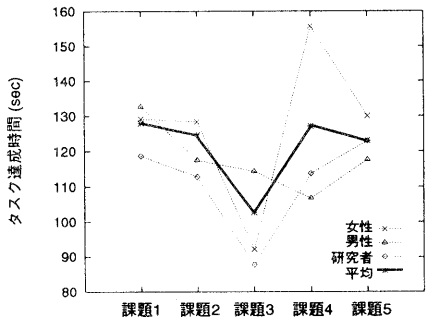

図 6: タスク達成時間

被験者全員の平均を見ると，音声入力中心の時 (課題 1，課題 2) のタスク達成時間は，キーボード入力や手書きの場合とほとんど変わらず，音声入力による時間短縮の効果は見られない．しかし，項目の移動などシステムの操作にマウスを併用できる課題 3 では，他の入力方法に比べて 20 秒以上時間が短縮されている．N-Best を利用できる課題 1 と，3 回目の発声までは N-Best を利用できない課題 2 では，N-Best を利用して発声回数を減らせるので課題 1 の方が早く作業が進むと予想していた．実際，発声回数は課題 2 の 2094 発声に対し課題 1 は 2020 発声で課題 2 の発声回数の方が多い．しかし，全被験者の平均タスク達成時間は課題 1 も課題 2 もほぼ等しく，男性や研究者は課題 2 の方が短時間で入力を終了している．これらの主な要因は，N-Best 候補から正解候補を選択するのに時間がかかっているためと推測できる．

マウスと音声入力の併用により，作業時間を大幅に短縮できることがわかった．また，現状では N-Best は候補の選択に時間がかかり，作業時間短縮には有効でないことがわかった．今後，N-Best の候補の用い方や表示方法を再検討する必要がある．

4.7 N-Best の利用

N-Best の利用に関するアンケート内容及び結果を図 7に示す．図の括弧内の数字は回答数である．

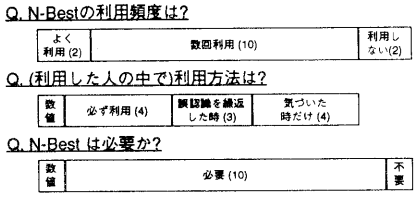

図 7: N-Best 利用に関するアンケート結果

14名中，12名の被験者がN-Best を利用している．利用しなかったと回答した被験者 2 名は，ともに認識率が高く，N-Best の利用を義務づけた課題 1 で 3 回利用したのみで課題 2 以降は実際に利用していない．また，このうちの 1 名は『発声し直した方が早い』という理由でN-Best は必要ないと回答している．利用方法に関する回答はまちまちだが，積極に利用しようとする被験者は 4 人で，残りの 8 人は，何度か誤認識を繰り返す中で利用している．N-Best の利用に何も制約してない課題 3 でのN-Best の利用回数は 72 回であった．全被験者で合計 130 回N-Best を利用する機会 (提示した 5 候補内に正解が含まれた回数) があったので，利用率は約 55% となる．また，認識率が高かった女性被験者で利用したのは 2 名のみで，合わせて 7 回だけである．しかし，最後の設問では 1 名を除いた全員がN-Best が必要と回答している．数値の入力のみ必要という意見もあった．これは実際の利用状況を反映した回答である．表 8にN-Best が利用された項目および回数を示す．

表 8: N-Best が利用された項目

項目	金額 (経路)	日付	時間	その他	合計
回数	63	23	39	32	157

N-Best 内に正解候補が表示されたのは全課題全被験者で合計 358 回で，44%にあたる 157 回N-Best が利用された．内訳を見ると，経路の項目の金額，日付，時間など数値の入力が約 80% を占めている．

以上のアンケート結果および実際のN-Best 利用状況から，N-Best は積極的に利用されるものではないが，データ入力用の音声入力システムでは必要であると考える．

5 システムの性能とユーザの使用感

テスト後に以下のアンケートを行なった．

[質問 1] 音声認識の性能は？

1———2———3———4———5

良い　　ふつう　　悪い

[回答] (1) 4 人，(2) 6 人，(3) 4 人，(4) 0 人，(5) 0 人．

[質問 2] 音声認識システムの使い勝手は？

1———2———3———4———5

使い易い　　ふつう　　使いにくい

[回答] (1) 3 人，(2) 7 人，(3) 2 人，(4) 2 人，(5) 0 人．

[質問 3] 音声入力するシステムについて

(1) このままでも是非使いたい

(2) こもまま使ってももいい

(3) 改善されれば使ってもいい

(4) 絶対に使いたくない

[回答] (1) 2 人，(2) 4 人，(3) 8 人，(4) 2 人，(5) 0 人．

[質問 4] 体感で何%くらい認識していると感じましたか？

[回答] (90%) 6 人，(85%) 1 人，(80%) 5 人，(70%) 2 人．

アンケート結果とシステムの性能を表 9に示す．表では，使い勝手のアンケート結果が同じ被験者の平均値を示してる．表中，’性能’ は質問 2 の回答，’利用’ は質問 3 の回答，’時間’ は音声入力時の平均タスク達成時間を表す．また，平均発声回数・平均操作回数はそれぞれ被験者が 1 データあたりに費やした発声回数・操作数 (発声数+マウス利用回数) の平均を表す．

被験者数が少ないため，表 9からは使い勝手に直接関わってくる性能がはっきりとは分からない．傾向として，被験者の印象である体感認識率が良いと，使い勝手がよく感じられるようである (図 8参照)．

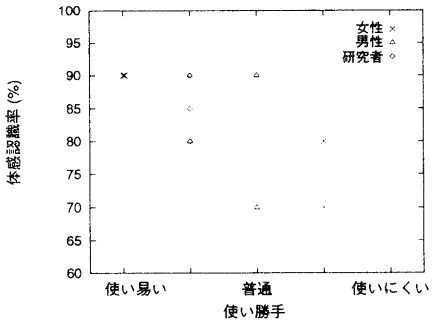

図 8: 体感認識率と使い勝手の相関

体感認識率と実際の認識率の相関を図 9に示す．

図中，横軸が被験者の体感認識率で，縦軸が実際の認識率である．また，図内の数字は，使い勝手に関するアンケートの回答である．斜線よりも上にプロットされた被験者は，実際の認識率よりも体感の認識率が悪いと感じた被験者である．使い勝手で評価が悪かった被験者は，実際の認識率と体感認識率の間の差が大きい．フィールドテストの記録から，これらの被験者に共通していたのは，誤認識した場合にデータ入力した発声がコマンドと誤認識された比率が高く，この誤認識により被験者が意図しない入力項目の移動が起こっている点で

表 9: アンケート結果とシステムの性能

使い勝手	認識率(%)	体感認識率(%)	性能	利用	時間(sec)	平均発声回数	アクセプト率(%)	平均操作回数
1	88.84	90.00	1.00	2.00	123.30	1.35	85.76	1.57
2	82.68	83.57	2.14	2.43	119.39	1.51	80.19	1.66
3	70.97	80.00	2.50	2.50	136.06	1.75	68.96	1.96
4	92.06	75.00	2.50	3.00	106.56	1.28	87.42	1.47

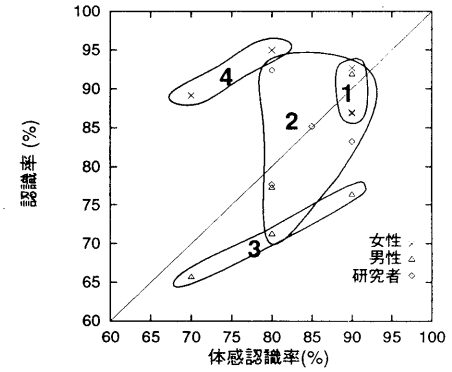

図 9: 体感認識率と認識率の相関

ある．また，システムが同じ誤認識を何度も繰り返している点も共通していた．これらの点が被験者に煩わしさを感じさせ，使い勝手の印象を悪くしたと推測される．一人の被験者はアンケートの感想欄に，「何度発声しても認識されないことがあって困った」と回答している．

使い勝手の評価を『4』(『使いにくい』と『普通』の間の評価) をした被験者を除くと，システムの使い勝手と性能に相関が見い出せる．認識率や平均発声回数，アクセプト率，操作回数とも使い勝手に影響していることがわかる．また，タスク達成時間は必ずしも使用感に影響するものでないこともわかる．システムの利用に関する回答との関係では，使い勝手が良いと回答した被験者ほど，本システムを実際に利用することに積極的である．

システムの性能として挙げた各項目は認識率を向上することによって，改善することができる．したがって，今後使い易いシステムの構築にあたり，基本的には認識率向上が必要と考えられる．システムの利用に関する設問で，『改善されれば利用する』と回答した被験者 8 名のうち，改善すべき事項を問う設問で『認識率の向上』を挙げた被験者が 6 名いた．しかし，前述したように認識率が高くても使い勝手が悪いと回答する被験者もいることから，ユーザの使用感に影響する誤認識した場合のシステムの挙動などユーザインタフェースの改良も重要な課題であると今回のテストであらためて判明した．

6 まとめ

音声入力機能を持つ社内の伝票を作成するためのデータ入力システムを試作し，被験者 14 名によるフィールドテストを行なった．その結果，(1) 適当な発声速度で入力した方が認識率が良い，(2) システム側の問題で入力時に約 4%のデータに欠損が生じる，(3)2〜3 回を入力を試みて認識されない語は，その後，発声を繰り返しても認識される可能性が少ない，(4) 現在の棄却処理の性能では入力項目に応じて使い分ける必要がある (5) 音声入力をマウスと併用して用いることにより，キーボード入力に比べ作業時間を短縮できる，(6) N-Best は作業時間短縮には寄与しない，(7) N-Best は積極的に利用されることはないがデータ入力システムには必要な機能である，(8) システムの使用感向上には認識性能の向上とともに誤認識時の対処が重要である，などの知見を得た．

これらの知見は今後のシステム構築指針として生かして行きたい．また，本報告で行なったフィールドテストでは被験者数が少なかったため，ユーザの使用感に関してシステムの性能との相関を見い出すのが難しかった．今後，テスト方法や調査項目を再検討し，被験者数を増やしたフィールドテストを再度行ないたい．

謝辞　本研究の機会を与えて頂いた，キヤノン (株) 情報メディア研究所 田村秀行所長に感謝致します．

参考文献

[1] 武田一哉: 音声認識 - フィールドテストの結果から見た問題点・研究課題 -, 信学技法, SP94-82, pp.25-30, 1995-1.

[2] 森島正俊, 磯部俊洋, 吉谷文徳, 小泉宣夫: ホームバンキングを想定した電話音声認識に関するアンケート, 日本音響学会講演論文集, 3-8-4, pp.153-154, 1996-3.

[3] 吉岡理, 荒井和博, 嵯峨山茂樹, 山田智一, 野田善昭, 井本貴之, 菅村昇: 音声入力機能を持つ住所入力システム, 日本音響学会講演論文集, 3-P-26, pp.213-214, 1995-3.

[4] 山田雅章, 山本寛樹, 小坂哲夫, 小森康弘, 大洞恭則: パラメータスカラ量子化と混合分布 HMM の次元独立演算による高速出力確率計算, 日本音響学会講演論文集, 2-2-16, pp.69-70, 1995-9.

[5] 小森康弘, 山田雅章, 山本寛樹, 小坂哲夫, 大洞恭則: N-Best 処理に向いた高速な未知語処理法の提案, 日本音響学会講演論文集, 2-5-7, pp.71-72, 1999-3.