

## 電話音声認識による資金移動サービスと対話の評価

磯部 俊洋<sup>†</sup> 森島 昌俊<sup>†</sup>  
吉谷 文徳<sup>†</sup> 小泉 宣夫<sup>†</sup>

近年、音声認識技術の発達により、システムがユーザの音声認識し、対話を行いながらサービスを行う音声対話システムが多数提案されている。我々は電話を用いた音声対話による資金移動サービスを想定し、その実現性について検討を行うために電話音声認識試作装置を開発した。電話音声認識試作装置は回線制御部を伴う音声認識ボードを6枚装着し、同時に6回線分の音声対話を処理することができる。本装置を使用し、ユーザの対話システムへの慣れと知識がシステムの性能に与える影響について測定した。評価は、90人の被験者に擬似的な振込サービスを行ってもらうことにより実施し、各被験者毎にシステムについての知識を与える前と与えた後の2回について、サービス達成率を測定した。その結果、ユーザが発声を繰り返すことによるシステムへの慣れと、システムについての知識の習得との双方について、それぞれサービス達成率の向上への効果がみられた。知識の与え方については、ユーザとシステムの対話例を被験者に試聴させることが分かった。評価試験の後に各被験者に対してアンケートを実施し、被験者が音声対話に対して感じる心理的な利便性について測定した。アンケート結果により約60%の被験者が電話による音声対話システムについて利便性を感じているとの回答を得た。

### Evaluation of Interaction in Voice Activated Phone Banking Service

TOSHIHIRO ISOBE,<sup>†</sup> MASATOSHI MORISHIMA,<sup>†</sup> FUMINORI YOSHITANI<sup>†</sup>  
and NOBUO KOIZUMI<sup>†</sup>

Recently, a development of speech recognition technique makes it possible to propose some speech dialogue systems that control interaction procedure with speech recognizer. In order to investigate the realization of the voice activated money transfer system, we have developed an experimental telephone speech recognition hardware which has six boards for dealing with simultaneous multi-channel speech dialogue processing. We have measured how the experience and the knowledge of users make the performance of the system higher using the speech recognition hardware. We have tested the voice activated telephone banking system with experimental money transfer dialogues collected from 90 subjects, and measured success rate of the service twice for each subject; one is done before showing them proper instructions and guidance and the other is after that. As a result, both of the experience and the knowledge of users made the success rates of the service higher, and showing examples of skilled user's speech dialogues was effective to get users accustomed to the system. We sent out a questionnaire after trials, and investigated how subjects felt about dialogue system with speech recognizer. As a result, 60% of them answered it was useful.

#### 1. まえがき

音声認識技術の発達は計算機等ハードウェアの処理能力の向上とともに大語彙連続音声認識のリアルタイム処理を可能にした。これにともないシステムがユーザの音声認識し、対話を行いながらサービスを行う音声対話システムの開発が進められるようになり、一

部については実用化が始まっている<sup>1)~3)</sup>。音声対話システムの有効な適用サービスの1つにコールセンター等の電話サービスがあげられる。欧米では、HMMベースの音声認識技術を用いてユーザの声を認識し、株価、映画や旅行等の電話による情報案内サービスの自動化が行われている<sup>4)~6)</sup>。

しかし、電話音声には雑音のほか、音声符号化、回線によるさまざまな歪みが混入されており、正確に認識を行うことはきわめて困難である。現在では電話音声の認識性能を高めるために、システムが主導をとり

<sup>†</sup> NTT データ通信株式会社情報科学研究所  
Laboratory for Information Technology, NTT DATA CORPORATION

ながらユーザに質問を提示し、ユーザの回答を予測可能な発話形態に限定したうえで、あらかじめ設定した文法に基いて受理する方法が主流となっている。また、ユーザが発声する「えーと」、「あー」のような不要語については、設定した文法の中に出現する可能性のある場所を仮定し、不要語対応の辞書を付加させることで受理する場合が多い<sup>7)</sup>。

文法的に不完全な自由発話音声認識し発話者の意図を自動獲得する技術は、研究レベルでは進展があるものの<sup>8)~10)</sup>、実用レベル、特に電話回線を経由する音声に対しては十分な性能を得られていない。自由発話音声データおよびテキストデータは収集が大変であり、統計的な処理に基づくボトムアップ的なシステム構築を困難なものとしている。したがって現在の電話音声認識では、ユーザは対話システムが設定する文法に沿って発声する必要があり、そのために、いかにユーザに対話システムについての知識を与えるかが重要になっている。また、ユーザが対話システムに対してどのような発声をするかを分析することは、今後自由発話音声認識し、ユーザの意図を獲得する対話システムを構築するためには必要なことであろう。

以上の理由により、我々は電話による資金移動サービスを想定し、その実現性について検討を行うために電話音声認識試作装置を開発した<sup>11)</sup>。本装置を使用した90名の被験者の模擬振込処理により、初めて音声対話を行うユーザの対話システムに対する発声の慣れ、知識がシステムの性能にどのような影響を与えるかについて評価および検討を行った。本システムは、最近サービスが始まりつつある金融機関のコールセンターサービスのオペレーションの自動化を目論んだものであり、電話で資金移動等のサービスが自動に行えるようにするものである。電話音声認識試作装置はハードウェアの制約により最大1,000単語までの同時認識が可能となっているが、認識語彙を変更したり、音声入力項目を階層化することで、その他のさまざまな電話を経由したサービスに適用することが可能である。本論文では開発した電話音声認識試作装置、資金移動アプリケーション、および試作装置を使用した音声対話の評価結果について述べる。

以下、2章では資金移動サービスにおいて本システムが取り扱う対話について、3章では電話音声認識試作装置について、4章では資金移動サービスにおける音声対話処理の評価結果について、5章では評価試験後に被験者に対して実施したアンケートの結果について述べる。

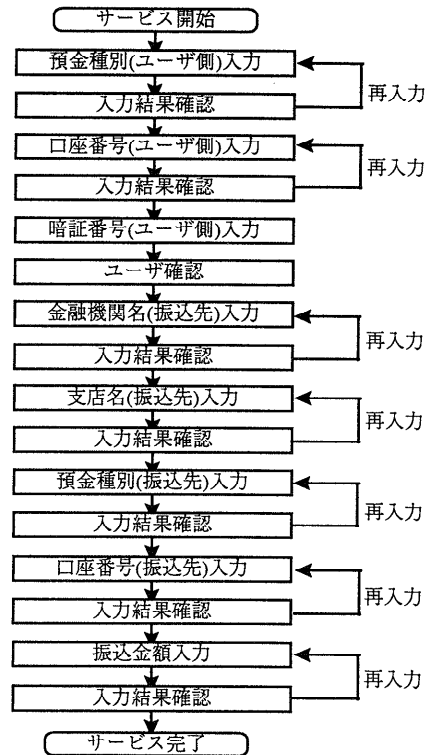


図1 資金移動サービスにおける音声対話処理

Fig. 1 Speech dialogue processing of money transfer.

## 2. 資金移動サービスにおける音声対話

資金移動サービスにおいてシステムがユーザの発声から獲得すべき情報は、ユーザの口座がある預金種別、口座番号、暗証番号と、振込先の金融機関名、支店名、預金種別、口座番号、振込金額の8項目である。図1に本システムが行う対話処理の流れを示す。

電話による音声対話であるため、各項目の入力後にシステムは認識結果を音声ガイダンスで出力し、認識結果が正解であったかを確認する。認識結果が誤りであった場合にシステムは再度同じ項目の発声をユーザに促す。実サービスにおいては、再入力の回数がある一定値を上回った場合にサービスは達成されずに中断されることになる。評価ではこの再発声の許容回数を変化させてサービス達成率を測定した。表1に音声対話例を示す。口座番号は7桁固定の連続数字認識であり、暗証番号は4桁固定の連続数字認識である。暗証番号を発声することはセキュリティを考慮するとプッシュボタン入力のほうが望ましい場合もあるが、音声対話の評価であることと将来的に話者照合技術の併用も考えられることにより音声入力とした。金融機関名は600の銀行名および信用金庫名が認識対象となつて

表 1 音声対話例

Table 1 Example of speech dialogue.	
「こちらは資本移動サービスセンターです。」	
「お客様の預金種別どうぞ。」	
	← 「普通。」
「普通でよろしいですか。」	
	← 「はい。」
「お客様の口座番号をどうぞ。」	
	← 「1234567。」
「1234567 でよろしいですか。」	
	← 「はい。」
「お客様の暗号番号をどうぞ。」	
	← 「1234。」
「お客様の振込処理を受け付けます。」	
「振込先の金融機関名をどうぞ。」	
	← 「データ銀行。」
「データ銀行でよろしいですか。」	
	← 「そうです。」
「振込先の支店名をどうぞ。」	
	← 「えー、京都です。」
「京都支店でよろしいですか。」	
	← 「はい。」
「振込先の預金種別をどうぞ。」	
	← 「普通です。」
「普通でよろしいですか。」	
	← 「はい。」
「振込先の口座番号をどうぞ。」	
	← 「9876543。」
「9876544 でよろしいですか。」	
	← 「違います。」
「もう一度振込先の口座番号をどうぞ。」	
	← 「9876543。」
「987653 でよろしいですか。」	
	← 「そうです。」
「振込金額どうぞ。」	
	← 「15,000 円です。」
「15,000 円でよろしいですか。」	
	← 「はい。」
「お客様の振込依頼を受け付けました。」	
振込内容を送信しますので、FAX を受信状態にしてください。ありがとうございました。」	

いる。支店名は、全金融機関共通に 10 店舗設定した。金額は桁読みで 1 億未満の金額を受け付ける。音声認識結果を確認するための発声は肯定用に「はい」、「そうです」の 2 単語を、否定用に「いいえ」、「ちがいます」の 2 単語を設定した。

### 3. 音声認識アルゴリズム

#### 3.1 音響分析

音声は、16 bit, 8kHz でサンプリングされた後、12ms ごとに 30ms のハミング窓によって切り出される。切り出された音声データは 0-11 次の LPC ケプストラム係数と 0-11 次のデルタケプストラムに変換され、これらの 24 次元のベクトルを音声特徴ベクトルとして使用した。

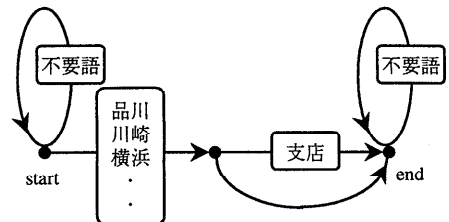


図 2 支店名認識用有限状態オートマトン  
Fig. 2 Finit state automaton for branch name recognition.

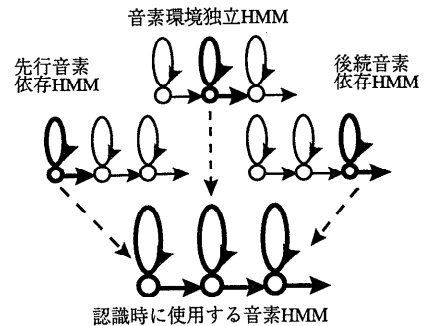


図 3 音素 HMM 構造  
Fig. 3 Composition of phone HMM.

#### 3.2 音声認識アルゴリズム

音声認識アルゴリズムは音素単位の HMM を音響モデルとして使用し、有限状態オートマトンにより文法が定義された連続音声認識である。音素 HMM は男性 200 名、女性 200 名の計 400 名から構成される実回線経由の電話音声データ<sup>12)</sup>から作成された 3 状態の left-to-right モデルであり、各状態あたり 10 個のガウス分布を持つ混合連続型である。認識処理には有限状態オートマトンによって文法が定義されるビタビウムサーチアルゴリズムを採用している。図 2 に支店名認識に使用した有限状態オートマトンを示す。「えーと」、「あのー」等の不要語は文頭において受理可能とした。

#### 3.3 音声認識アルゴリズムの高速化

音声対話に用いる音声認識処理はリアルタイム処理が前提条件となる。音素 HMM は前後音素環境に依存したコンテキスト依存モデルを基本としているが、認識時には演算量、使用メモリを低減するため、図 3 に示すように先行音素依存 HMM の第 1 状態、音素環境独立 HMM の第 2 状態、および後続音素依存 HMM の第 3 状態から前後音素環境依存 HMM を合成した。この結果、通常の前後音素環境依存モデルを使用した場合の HMM の総状態数は 2,100 状態であったのが、本方式により認識性能の劣化を招くことなく 500 状態

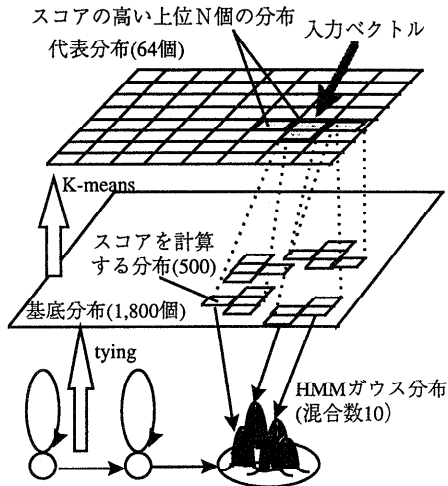


図 4 HMM の出力確率演算

Fig. 4 Calculation of output probability of HMM.

へ低減することができた。

混合連続型 HMM は認識処理のうち出力確率演算の計算量が多いことが問題となる。これを回避するため、図 4 に示すように各音素 HMM のガウス分布を異音モデルごとに共有し、合計 1,800 個の基底ガウス分布を作成した後、さらに K-mean 法によりクラスタリングをして 64 個の代表ガウス分布を作成した。認識時には入力音声のフレームごとにまず 64 個の代表ガウス分布のスコアを計算し、基底ガウス分布についてはスコアの高い  $N$  個の代表ガウス分布に属するものに限定して計算を行い、それ以外の基底分布は代表分布のスコアを代用した。 $N$  の値は、演算する基底ガウス分布数が 500 になる値とした。この結果、スコアを計算するガウス分布数は、5,000 個から 564 個へ大幅に削減することができた。

4. ハードウェア構成

図 5 に開発した電話音声認識試作装置の構成図を示す。本装置はワークステーションであり、回線制御と音声認識を行うボードを 6 枚搭載している。このため 1 台の装置で同時に 6 回線分の音声対話を扱うことが可能である。音声対話管理はワークステーションが行う。図 6 に音声認識ボードの構成図を示す。1 枚のボードには回線制御部とアナログ/デジタル変換器の他、DSP3 個と MPU1 個を搭載する。各 DSP および MPU の行う処理を表 2 に示す。

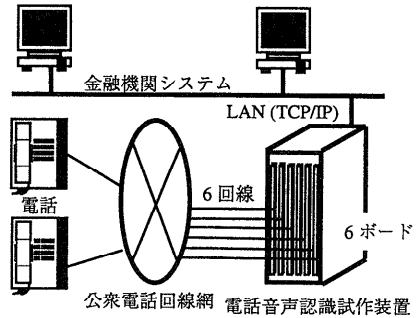


図 5 電話音声認識試作装置

Fig. 5 Experimental telephone speech recognition unit.

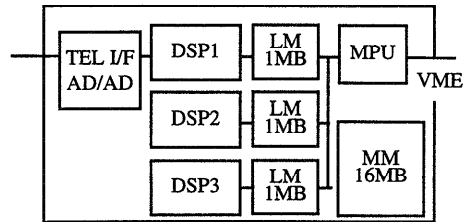


図 6 音声認識ボード

Fig. 6 Speech recognition board.

表 2 DSP および MPU の処理  
Table 2 Process of DSP and MPU.

DSP 1	LPC ケプストラム算出 デルタケプストラム算出 基底分布出力確率算出
DSP 2	HMM 出力確率算出
DSP 3	ビタビウムサーチ
MPU	バックトレース DSP 制御

5. 電話音声認識試作装置による音声対話の性能評価

5.1 電話音声認識試作装置の基本性能評価

実回線を経由した音声対話の評価の前に電話音声認識試作装置の基本性能について測定した。擬似交換機経由で本装置へ音声データを入力し、7桁固定連続数字、金融機関名、金額の認識率について測定した。評価データは話者 400 人の音声データであり、学習データに対してオープンである。各タスクの語彙サイズ、パープレキシティ、認識率を表 3 に示す。金額認識のパープレキシティが 24.3 であり、最も認識が困難であり、認識率は 77.4%であった。

5.2 資金移動サービスにおける音声対話評価

電話音声認識試作装置を用いた資金移動サービスにおける音声対話について評価した。実験では、被験者が初めて音声対話を行う場合と、音声対話を 1 回経

表 3 音声認識装置の基本性能測定結果

Table 3 Results of bench mark test of speech recognition hardware.

タスク	語彙サイズ	Perplexity	認識率
7桁連続数字	12	12.0	92.4%
金融機関名	600	4.94	90.8%
金額	46	24.3	77.4%

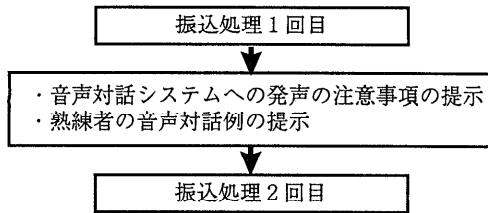


図 7 各被験者への評価手順

Fig. 7 Filed trial sequence for each subject.

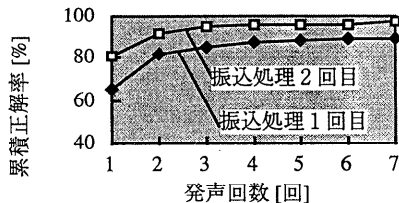


図 8 発声回数に対する 7 桁数字累積認識率

Fig. 8 Cumulative accuracy rate vs. number of repeat attempts in recognizing seven connected digits.

験し、システムについての知識を与えられた場合の音声対話システムの性能に与える影響について比較を行った。

実験は、90人の被験者に擬似的な振込処理を行ってもらうことにより実施した。90名の被験者の構成は男性55名、女性35名であり、すべて音声認識を用いた対話を行うのは初めてである。振込内容については評価前に各被験者ごとに内容の異なる振込依頼書を被験者に配布しておいた。振込処理は各被験者ごとに2回実施し、1回目と2回目の間にシステムへの発声の仕方についての注意事項と、熟練者の音声対話例を被験者に聞かせた(図7参照)。1回目とシステムへの発声の仕方についての注意事項、および熟練者の音声対話例の提示、2回目は時間を空けずに連続して行った。発声方法についての音声ガイダンスでは、1) 不要語はなるべく発声しないこと、2) 数字の4は「よん」、7は「なな」と発声すること、3) システムのガイダンスが終了してから発声することの3項目について被験者に提示している。

1回目の振込処理と、2回目の振込処理の、7桁連

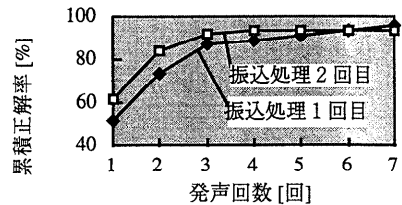


図 9 発声回数に対する金融機関名累積認識率

Fig. 9 Cumulative accuracy rate vs. number of repeat attempts in recognizing bank name.

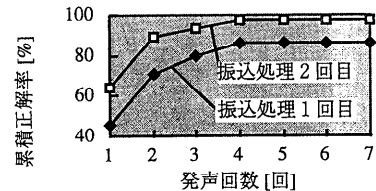


図 10 発声回数に対する金額累積認識率

Fig. 10 Cumulative accuracy rate vs. number of repeat attempts in recognizing money amount.

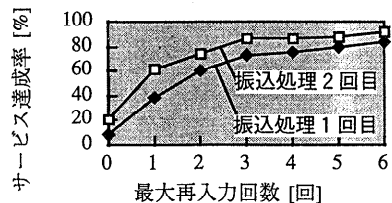


図 11 最大再入力回数に対するサービス達成率

Fig. 11 Success rate of money transfer service vs. number of repeat attempts in each recognition task.

続数字認識率、金融機関名認識率、金額認識率を図8、図9、図10に示す。横軸は繰返し発声回数であり、縦軸は繰返し発声回数に対する累積認識率を示している。全タスクの1回目、2回目ともに繰返し発声することにより認識率は減少しているが、発声回数が3~4回程度で収束している。これは被験者が発声を繰り返すことで音声対話システムへの発声に慣れたためといえる。1回目と2回目を比較すると、発声の仕方や音声対話例を示した後の2回目のほうが10%から20%認識率が上がり、金融機関名と比較して発話方法のバリエーションが多い7桁連続数字および金額の認識率の向上した度合いが大きい。また、7桁連続数字および金額の収束した後の認識率が1回目より2回目のほうが高いことから、被験者がシステムへの発声の慣れだけでなくシステムについての知識を持っていることが重要であることが分かる。

1回目の振込処理と、2回目の振込処理のサービス達成率を図11に示す。サービス達成率は、(振込処

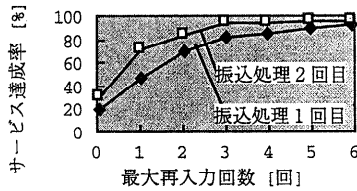


図 12 連続数字認識、金額認識をプッシュボタンに置き換えた場合の最大再入力回数に対するサービス達成率

Fig. 12 Success rate of money transfer service vs. number of repeat attempts in each recognition task with tone detection for the connected digits and money recognition.

理が正常に完了した回数)/(評価回数) × 100 で表される。横軸は各入力項目ごとに許容させる再入力回数であり、本評価では、最大 6 回を設定した。たとえば、許容再入力回数が 3 の場合、各入力項目別に 3 回再入力しても正しく認識されなかったらサービス中止となり、サービスは達成されなかったこととなる。縦軸はサービス達成率を表す。本評価では、1 回目の振込処理、2 回目の振込処理ともに許容再入力回数が 3 回まではサービス達成率が上昇し、3 回以降、収束している。また、2 回目の振込処理のほうが 1 回目より 10% から 20% サービス達成率が高く、2 回目の振込処理においてサービス達成率が収束し始める再入力回数 3 のとき、約 85% のサービス達成率であった。収束後のサービス達成率が 1 回目より 2 回目のほうが高いことから、サービス達成率に対しても被験者がシステムへの発声の慣れだけでなく、システムについての知識を持っていることが重要であることが分かる。

サービス達成率が 85% であることから、音声入力のみでサービスを実運用するには十分とはいえない。そこで、図 1 に示す音声対話処理において口座番号等の連続数字および振込金額の入力をプッシュボタンに置き換えた場合のサービス達成率を本評価結果より推定した。システムのプッシュボタンの認識率は 100% とした。推定したサービス達成率を図 12 に示す。音声入力のみによるサービス達成率と同様に振込処理 1 回目、2 回目ともに許容再入力回数が 3 回まではサービス達成率が上昇し、3 回以降、収束している。振込処理 2 回目において、サービス達成率が収束し始める再入力回数 3 のとき、約 95% のサービス達成率であった。金融機関名については、ユーザが簡単に機関コード番号を思い起こすことが困難であることを考えると、入力項目によって音声入力のほうが望ましい場合もある。音声認識の性能からみた場合においても、音声入力とプッシュボタン入力を併用すれば、よ

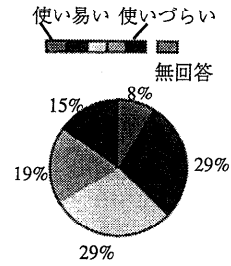


図 13 音声認識は使いやすいか  
Fig. 13 Is speech recognition useful?

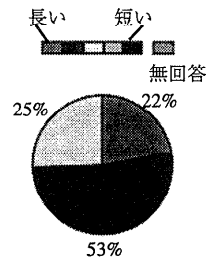


図 14 サービスにかかる時間は長いか短いか  
Fig. 14 How about the duration of speech recognition service?

りサービス達成率を向上させた音声対話システムの構築が可能といえる。

### 6. アンケート調査

音声対話システムのマン・マシンインタフェースについての利便性は、ユーザがシステムによるサービスを受ける際にどのように感じるかによるところが大きい。ユーザがシステムに対して感じる利便性を調べるため、資金移動サービスにおける音声対話評価の被験者 90 人に評価後、アンケート調査を実施した。アンケートでは、各質問項目に対して 5 段階評価で回答を設定した。各アンケート項目に対する回答結果を図 13 から図 20 に示す。

図 13 は、音声認識は使いやすいか使いづらいかという質問項目に対する回答である。約 3 分の 1 である 37% の被験者がおおむね使いやすいと回答し、34% の被験者が使いづらいと回答し、残り 29% の被験者がどちらともいえないと回答した。

図 14 は、音声対話によるサービスにかかる時間についての質問に対する回答である。75% の被験者が長い、あるいはどちらかといえば長いと回答した。資金移動サービスの各入力項目ごとに認識結果に対して確認を行っているため、ATM のディスプレイ表示に比べて時間を要する音声ガイダンスの出力回数が増加し、

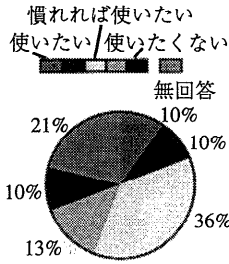


図 15 実サービスで音声認識を使いたいか  
Fig. 15 Are you willing to use speech recognition as an actual service?

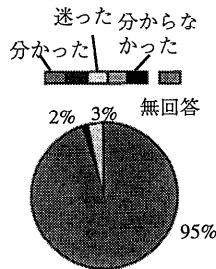


図 16 口座番号を聞かれたとき発音方法が分かったか  
Fig. 16 Is it easy to think of what to say when you are asked your account number?

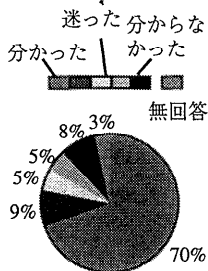


図 17 振込金額を聞かれたとき発音方法が分かったか  
Fig. 17 Is it easy to think of what to say when you are asked money amount?

サービス全体にかかる時間が長くなってしまったためと考えられる。これは約3分の1の被験者が音声認識は使いづらいと回答した要因の1つといえる。特に熟練したユーザに対しては、ガイダンスを最後まで聴くことは煩わしく、システムのガイダンスを中断して音声入力を開始できる機能は必要不可欠であるといえよう。

実サービスに音声入力を使用したいかという質問に対しては、音声対話に慣れた状態ならという条件であれば56%の被験者が実際に使ってみたいと回答している(図15参照)。

図16と図17は、音声入力する項目に対する発話

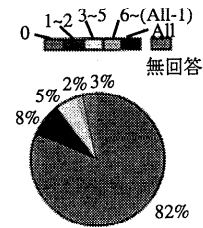


図 18 不要語を何回発音したか  
Fig. 18 How many times did you say words not in the vocabulary?

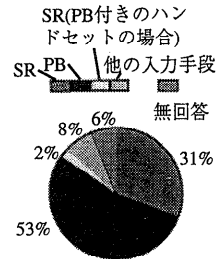


図 19 口座番号や金額を入力するのに SR と PB のどちらを使用したいか  
Fig. 19 Do you prefer SR or PB for inputting account number and money amount?

方法のバリエーションの多さが被験者にどのような影響を及ぼすか調査したアンケート結果である。口座番号のように発声の方法が連続数字であると分かりやすい入力項目は95%の被験者が迷うことなく発声を開始している。発声の方法が、連続数字と桁読みの2通りあると考えられる金額では、その比率は70%まで低減している。プッシュボタン入力による各種サービスの普及により、1回目の振込処理では、音声入力ではなく、プッシュボタンを押す被験者も数名いた。音声認識装置では、考えられる発声方法をすべて受理するように有限状態オートマトンを設定するか、発声方法についてユーザに知識を与える必要があるといえる。

図18は1回目の振込処理のときに何回不要語を発声したかについての回答である。90%の被験者が不要語の発声回数が1以下であり、相手がシステムであると分かっている場合には、人は不要語をあまり発声しないことが分かる。

図19は口座番号や振込金額等、入力項目を容易に数字で表すことができる場合に、音声入力とプッシュボタン入力ではどちらを使いたいという質問に対する回答である。プッシュボタン入力の普及により、音声入力を選択した31%の被験者に対して53%の被験者がプッシュボタン入力を選択した。音声入力やプッシュボタン入力等のモダリティは入力項目によって使

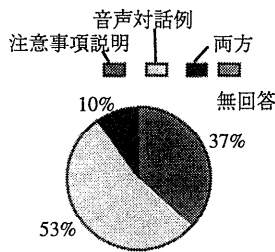


図 20 注意事項説明と音声対話例ではどちらが役にたったか  
Fig. 20 Which is better for support explanation or example dialogue?

い分けたり、ユーザの好みでどちらでも受け付ける必要があるといえる。同質問で2%の被験者が電話器のハンドセット側にプッシュボタンが設置されている場合に限って音声入力を選択しているように、モダリティは、ユーザの使用する端末によって切り替える必要もあろう。

図 20 は 1 回目の振込処理と 2 回目の振込処理の間に被験者に対して聴かせたシステムへの発声方法のガイダンスと熟練者の対話例のどちらが役にたったかという質問に対する回答である。37%の被験者がガイダンスを選択したのに対し、53%の被験者が音声対話例を選択した。熟練したユーザの対話例を聴くことが音声対話についての知識をユーザに与えるのに効果的であるといえよう。

## 7. む す び

電話音声認識試作装置と、それを用いて行った電話上での音声対話による資金移動サービスの評価および評価後に実施した被験者へのアンケート結果について述べた。被験者の音声対話システムへの発声の慣れと音声対話システムについての知識はともにシステムの性能を向上させるのに効果が認められた。被験者が音声対話を 1 回経験し、システムについての知識を与えられた後のサービス達成率は、最大再入力回数を 3 にした設定した場合に約 85%であった。また連続数字入力と金額入力をプッシュボタン入力にした場合のサービス達成率は約 95%まで向上した。完全な自由発話音声認識する技術が不十分な現在では、いかにユーザに音声対話システムについての知識を与えるかが重要といえる。被験者に対するシステムの知識の与え方では、発声方法についてのガイダンスよりも音声対話例を好む被験者のほうが多かった。しかしアンケートでは多くの被験者が音声対話ではサービスを完了するまでの時間が長いと回答しており、再入力回数は 3 回程度が限度といえる。

## 参 考 文 献

- 1) 竹林洋一, 金澤博史, 永田仁史, 瀬戸重宣: 不特定話者音声自由対話システム TOSBURG—マルチモーダル応答と音声応答キャンセルの利用, 信学論 (A), Vol.J77-A, No.2, pp.241-250 (1994).
- 2) 森屋裕治, 阿部野尚, 山本幹男, 中川聖一: 対話予測を利用した音声による観光案内システム, 信学技報, SP92-109 (1992).
- 3) Kuroiwa, S., Sakayori, S. and Yamamoto, S.: Prank Call Rejection System for Home Country Direct Service, *Proc. IVTTA'96*, pp.135-138 (1996).
- 4) Aktas, A.C. and Hoge, H.: Voice Recognition in Siemens Private and Public Telecommunication Systems, *Proc. AVIOS'96*, pp.133-141 (1996).
- 5) Lennig, M., Sharp, D., Kenny, P., Gupta, V. and Precoda, K.: Flexible Vocabulary Recognition of Speech, *Proc. ICSLP'92*, pp.93-96 (1992).
- 6) Bennacef, S., Devillers, L., Rosset, S. and Lamel, L.: Dialog in the RAILTEL Telephone-Based System, *Proc. ICSLP'96*, pp.550-553 (1996).
- 7) 井上直己, 武田一哉, 山本誠一: ガーベジ HMM を用いた自由発話文中の不要語処理手法, 信学論 (A), Vol.J77-A, No.2, pp.215-221 (1994).
- 8) Riccardi, G., Gorin, A., Ljolje, A. and Riley, M.: A Spoken Language System for Automated Call Routing, *Proc. ICASSP'97*, pp.1143-1146 (1997).
- 9) Albesano, D., Baggia, P., Danieli, M., Gemello, R., Gerbino, E. and Rullent, C.: Dialogos: A Robust System for Human-Machine Spoken Dialog on The Telephone, *Proc. ICASSP'97*, pp.1147-1150 (1997).
- 10) Kawahara, T., Lee, C. and Juang, B.: Combining Key-Phrase Detection and Subword-Based Verification for Flexible Speech Understanding, *Proc. ICASSP'97*, pp.1159-1162 (1997).
- 11) Isobe, T., Morishima, M., Yoshitani, F., Koizumi, N. and Murakami, K.: Voice-Activated Home Banking System, *Proc. ASR Workshop*, pp.163-164 (1995).
- 12) 森島昌俊, 磯部俊洋, 村上憲也: 電話音声データベースの構築と評価, 音講論, 2-8-8, pp.53-54 (1994).

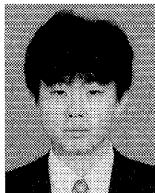
(平成 9 年 6 月 30 日受付)

(平成 10 年 3 月 6 日採録)

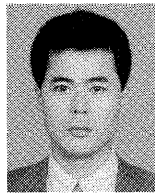


**磯部 俊洋**

1990年名古屋大学工学部電気系工学科卒業。同年NTTデータ通信(株)に入社、電話音声認識装置の開発を行う。現在技術開発本部情報科学研究所ヒューマンコミュニケーション研究担当において音声認識技術の研究開発に従事。日本音響学会会員。

**森島 昌俊**

1993年名古屋大学工学部電気系工学科卒業。同年NTTデータ通信(株)に入社、電話における音声対話の研究を行う。現在技術開発本部情報科学研究所ヒューマンコミュニケーション研究担当において音声認識技術の研究開発に従事。日本音響学会会員。

**吉谷 文徳 (正会員)**

1994年徳島大学工学部情報工学科卒業。1986年同大学院修士課程修了。同年NTT入社。現在NTTデータ通信(株)技術開発本部情報科学研究所ヒューマンコミュニケーション研究担当において主に、データ通信システム、音声認識/合成システムの開発に従事。

**小泉 宣夫 (正会員)**

1975年京都大学大学院工学研究科修士課程修了(精密工学専攻)。同年日本電信電話公社(現NTT)電気通信研究所に入社、音響振動、拡声通信機器の研究開発を行う。引き続き、ヒューマンインタフェース研究所にて音場創成の研究に従事。この間(1988~1989年)マサチューセッツ工科大学機械工学科客員研究員。現在NTTデータ通信(株)情報科学研究所において音声処理技術、知識処理技術の研究開発を担当。工学博士。