

解説

— 音声認識技術への応用(2) —

バンキング音声応答認識システム†

八橋主計^{††} 石井直樹^{†††}

1. はじめに

バンキング・システムでは総合オンライン化が進められており、CD・ATMなどの自動窓口装置の普及により、営業店業務の省力化と顧客サービスの向上がなされている。しかし、振込通知や残高照会などの営業店業務の機械化が取残されており、その効率化が強く望まれている。

最近、プッシュホンを端末として電話機から押しボタンダイアル信号の形で情報を受信し、情報処理装置の処理結果を音声の形に変換して電話機に情報出力するシステムが出現しており[†]、振込通知などのバンキング・サービスに使われ始めている。この種のサービスを、広く普及している回転ダイアル電話機の加入者に対して実施したいという要求が高まっている。

プッシュホンの場合には、回線接続ののち端末から押しボタン信号の形で音声以外の信号を送出できるので、上述のようなシステムが可能になる。しかし、回転ダイアル電話機の場合には、回線接続の後は音声以外の送出手はできないので、直接音声の形で情報を投入する方法が望まれる。すなわち、情報処理装置の入口に音声認識装置を備え、端末からの音声信号を認識し、受信情報を情報処理装置に伝える方式である。

音声認識装置は、従来商品仕分けなどの物流制御の分野で主に使用されていたが、このような通信の分野にも適用領域を拡げつつある。電話機を端末として音声で情報処理システムと対話する方式は、電話の普及性、サービスの簡便性の点で今後の発展が予想される。その典型的な例がバンキング・サービスへの適用である。

そこで本稿では、まず通知・照会など電話サービス

関係の銀行業務の内容を概観し、ついでそれを自動化するのに必須となる音声認識技術、および音声認識装置とシステムの構成・処理方式について述べ、最後に今後のバンキング音声応答認識システムの発展形態について展望することとする。

2. 銀行における通知・照会関連業務の概要

種々の銀行業務のうち、電話を用いて実施可能なサービスは、表-1に示すように多種存在する。これらの業務の中で、特に業務量の多いのが振込通知である。ある銀行の本店所在地域における年間の業務量は、振込通知が約4百万件であり、照会業務はその3割程度となっている。

通知業務の場合には、銀行センタからの通知データの作表結果をもとに、扱者が順次顧客に電話をかけ通知内容を知らせる作業になっている。照会業務の場合には顧客からの電話依頼に応じて、扱者が端末から口座番号等の必要情報を投入してセンタから情報を引出し顧客に伝えている。これまでは、一部テレックスによる自動振込通知が実施されているほかは、上記のように人手による処理が行われている。従来の処理法の問題点は、

- i) 業務量の増大に対処しにくい、
 - ii) 人件費の高騰により、経費がかかる、
 - iii) 通知に際し時間遅れが生じる、
 - iv) 誤りが発生する可能性がある、
- などである。

上述の業務内容から推察されるように、センタと顧客との間でやりとりする情報は決りきったものである

表-1 電話サービス関連の銀行業務

通知サービス	照会サービス	依頼サービス
振込通知	振込照会	振替依頼
引落とし通知	残高照会	振込依頼
資金不足通知	引落とし照会	ローン予約
満期通知	ローン案内	
	外貨レート案内	

† Audio Response and Recognition System for Banking Service by Kazuo YATSUHASHI (Data Communications Bureau, N. T. T.) and Naoki ISHII (Yokosuka Electrical Communication Laboratory, N. T. T.).

†† 日本電信電話公社データ通信本部

††† 日本電信電話公社横須賀電気通信研究所

ので、そのような業務を自動化することが可能になる。その方法として、電話網と銀行センタを音声応答認識システムを介して接続し、顧客が直接センタと音声でやり取りする方法が有望となる。このようなサービスのねらいは、省力化、効率化、迅速化、正確さ、使いやすさにある。

3. 音声認識技術

3.1 従来の適用分野および使用法

現状の音声認識技術は、任意の人が自由に話した言葉を理解できるレベルには到達しておらず、種々の制約や約束の下に限定された範囲の言葉を識別できない。主な制約条件は下記の通りである。

i) 使用者はあらかじめ自分の声を装置に登録せねばならない。(特定話者音声認識)

ii) 認識対象の言葉は、数十～数百個の単語に限られる。(限定単語認識)

iii) 発声は、通常単語ごとに区切りを入れる形式である。(離散発声)

これまでの音声認識装置の主な適用領域は物流制御の分野であり、その場合には定まった範囲の使用者が大量の情報投入を行うので、話者の登録(学習)が可能であり、登録済の話者が登録済の単語を入力するにはさして問題にならない。

音声認識装置の適用分野を拡大し、かつさらに使いやすいものにするために、上記の制約を緩和することが必要である。そのために種々の検討がなされており、連続発声単語を認識する装置、不特定話者の単語音声認識する装置などが最近開発されている²⁾⁻⁴⁾。

3.2 電話入力音声の認識

情報処理装置への入力手段として電話機からの音声を用いることは、電話の普及性・簡便性の点から優れていることは先に述べた。しかし、その実現には、下記のいくつかの項目を解決しておかねばならない。

(1) 電話系を経由することによる影響: 入力音声は電話系の特性および周囲環境の影響を受けている。音声認識に影響を与えるこれらの要因として次のものが挙げられる。

i) 周波数帯域がほぼ 0.3~3.4 kHz に制限されており、かつ回線により遮断特性に若干の差がある。

ii) 使用する電話機の種類(4号, 600形, 601形, 701形), および加入者線の線種(0.32~0.9 mmφ)と線長により周波数特性が異なる。

iii) 到達する信号のレベル変化範囲が 20 dB 以上と

大きい。

iv) 通常の電話機は炭素送話器を用いており、それに基づく非線形歪がある。

v) SN 比が十分でなく(30 dB 以下), かつ重畳している雑音の種類もさまざまである。

(2) 不特定話者の音声認識: 電話を用いて広くサービスを実施する場合、利用者は不特定多数となるので、あらかじめ利用者の音声を学習し装置に登録しておくことはできない。各利用者がデータ投入前に本人の声を装置に学習させることも不可能ではないが、特に投入すべきデータ量が少ない場合には、学習のための時間が長くなりサービス性を低下させるので好ましくない。すなわち、あらかじめ話者の学習を行わない音声認識—不特定話者音声認識—の技術を用いることが必要になる。話者が不特定であることの技術的問題点をまとめると次のようになる。

i) 話者により長時間スペクトル構造が異なり、かつ音素ごとのスペクトル構造もばらつきを示す。

ii) 発声音量がさまざまであり、送話器の非線形性により周波数特性も影響を受ける。

iii) 単語などの発声に際し、構成音素の継続時間が話者によりさまざまである。

iv) 単語の構成音素そのものが、無声化の有無などの点で話者により異なる。

これらの点を考慮した不特定話者の音声認識技術については次節で詳述する。

(3) 認識対象語の選定: 特定話者の音声認識では、学習時に使用者のほかに認識対象語を登録することになる。不特定話者音声認識の場合には、サービスを実施する上で必要となる認識対象語をあらかじめ定めて、その語のみを認識するように装置を構成せねばならない。認識対象語の変更・追加は、特定話者の場合に比して困難になるので、システム構築前に認識対象語を吟味・選定することが重要となる。

3.3 電話系・不特定話者向けの音声認識法

上述のように、電話系・不特定話者を対象とする音声認識では、諸特性の変動を吸収できる認識法が鍵となる。ここでは、そのような効果が期待できる認識法と処理法について述べる。

(1) 適応的学習: 特定話者の音声認識では 99% 程度の高い認識性能が得られていることに着目し、サービス中の音声を学習に用いて後からの発声の認識に役立てる方法が考えられる。電話機から投入するデータ量が多い場合には、適応的学習の効果が収束し、

それ以降の入力データには高い認識性能を期待できる。しかし、データ量が少ない場合には、学習に用いるデータが限られあまり大きい効果を期待できない。

(2) 話者等による特徴量変化を正規化する方法：電話系不特定話者の音声認識では、話者等の変化に独立な音声特徴量を用いることが望ましい。そのような特徴量を探索する試みとしてホルマント周波数の変換などの方法があるが、まだ決定的なものには至っていない。また、音声特徴量の完全なる正規化ではないが、話者固有のスペクトルを含めて伝送系の周波数特性を適応逆フィルタで等化する方法がある⁵⁾。勿論この方法だけでは十分ではないが、不特定話者向けの認識前処理としてかなりの効果がある。そのほか、話者等による特徴量変化を平均化した標準パターンを用いる方法がある。この方法は、特定話者用の認識装置に平均化した標準パターンを登録することにより不特定話者用に使用できるという簡便性があり、かつある程度効果もあるが、不特定話者に対して高い認識性能を期待するまでにはゆかない。

(3) 話者等による特徴量のばらつきを見込んだ認識法：話者等による特徴量のばらつきを分析し、その結果を積極的に利用して高い認識性能をねらう2つの方法が開発されている。そのひとつは、パターン・マッチング法において、特徴量の分布から求めた代表的な複数の標準パターンを準備しておき、入力パターンと動的計画法(DP)に従ってマッチングを行う方法である。ほかの方法は、認識対象語を互いに区別できる識別関数の特徴量の分布から求めておき、入力音声に対し特徴空間上の位置計算を行って認識する方法である⁶⁾。いずれの方法においても、電話系を通して不特定多数が発声した認識対象語の特徴量分布を正確に把握することが肝要であり、それが認識性能を支配するものといえよう。

3.4 電話入力用音声認識装置の構成

電話入力用音声認識装置の構成として、複数の標準パターンを内蔵するパターン・マッチング法を例にして説明する。なお、音声分析は最尤スペクトル推定に基づく方法で、単語の認識法として、入力音声と音素標準パターンとの類似度系列を音素系列で表現された単語辞書とDPマッチングする方法を採り上げる⁶⁾。

図-1に装置の概略構成を示す。図-1において音声入力部は、増幅器、低域濾波器、AD変換器などから構成されており、入力音声をデジタル信号に変換する。音声分析部では、15ミリ秒ごとの音声に対し特徴

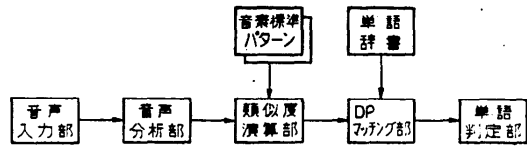


図-1 音声認識装置の構成

量として最尤スペクトル・パラメータを算出する。また、スペクトル平坦化などの前処理、および入力信号中の音声部分の切出し処理も音声分析部で行われる。

類似度演算部では、音声分析部から出力される特徴量の時系列に対し、あらかじめ登録されている音素標準パターンを用いて、各音素との類似度の時系列が計算される。音素標準パターンは、認識対象語を構成する各音素(2重母音の過渡部などを含む)の音声特徴量の典型的な値を登録したものである。前述のように、この方法では電話系の特性や話者による声質の相違を考慮して、各音素に対し複数の標準パターンを登録しておく。標準パターンの作成は、種々の条件下で多数の話者が発声した認識対象語の音声データを分析し、各音素ごとに音声特徴量をまとめ、統計的手法により類別化して選定される。実験的検討によれば、用意すべき標準パターンの組数は8~16個で十分である。

このようにして、類似度演算部から入力音声に対する音素類似度系列が出力される。DPマッチング部では、この系列と単語辞書に書かれている音素系列とのマッチングが行われ、入力音声と登録単語との類似度が計算される。単語辞書には、認識すべき単語が音素記号系列の形で登録されており、無声化の有無、長音の発声の仕方の相違などを考慮して、ひとつの単語に複数の音素系列を与えている場合もある。

単語判定部では、DPマッチング部から送られてくる単語類似度の中から値の大きな単語を選び出し、類似度の値とともに上位の処理装置に出力する。上位の処理装置では、この認識結果と利用者との会話状況から最終的に入力単語を決定する。なお、最大の類似度がある閾値に達していない場合には棄却の判定がなされる。

電話入力用の音声認識装置では、同時に多数の利用者の音声入力が可能のように、図-1の構成が複数台実装される。その際、音声入力部と音声分析部の間に集線のためのリンク部を設け、音声分析部以降を時分割的に利用する工夫もなされている。

ここでは、音素を単位として標準パターンを登録し

ておくパターン・マッチング法の装置構成と動作原理を述べたが、単語を単位として標準パターンを登録しておく方法もある。音素を単位とする方法は、日本語を構成する各音素の標準パターンを準備しておけば、音素系列を単語辞書に登録するだけで、認識単語の追加・変更ができるという特長がある。

4. システム構成および処理方式

4.1 システム構成

通知・照会関係の銀行業務を、音声応答と音声認識を用いて実施するシステムの構成、およびその実現に必要な装置類について述べよう。図-2に音声応答認識システムの構成例を示す。

音声応答認識システムは、片や加入電話回線を通して銀行の顧客の電話機と、片や中継用電算機などを經由して銀行のオンライン・センタと接続されている。中継用電算機は、銀行センタと音声応答認識システムとの間の電文の管理などを行うもので、センタから送られてくる振込通知データを格納する振込通知ファイルと、顧客の加入者番号・暗証番号などを格納した顧客登録ファイルを有する。音声応答装置、音声認識装置、回線対応装置、制御装置から構成されている。

音声応答装置は、発着信時の応待、顧客へのガイダンス、通知あるいは照会内容などを音声の形で送出するための装置である。音声応答装置は、このような内容の音声を送出するために、多種類の応答文や単語を生成するほか、通常センタから仮名文字の形式で送られてくる人名や会社名を音声に変換する任意語生成機能を具備せねばならない。音声認識装置は、顧客が発声した単語を識別し、結果を制御装置に伝える。

回線対応装置は、双方向トランク、自動着信部、自動発信部などからなり、電話網とのインタフェースの役目を果たす。照会業務で顧客が発呼した場合は、自動着信部がそれを検知し制御装置に伝える。通知業務の場合には、センタからの通知データに基づき制御装置が自動発信部に指示を出し、顧客が応答するとそれが制御装置に伝えられ、制御装置は音声出力など所定の処理を進める。双方向トランクにおける音声送受のための切替も制御装置の指示の下に行われる。

制御装置は、中継用電算機とのデータ送受信、各装置の制御のほか、受信電文の分析、音声応答文の作成、文脈を考慮した認識結果の判定、入力データの分析・

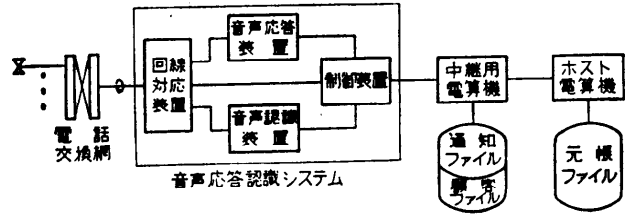


図-2 音声応答認識システムの構成例

チェックなどの処理をプログラム制御の下に行う。

4.2 処理方式

バンキング音声応答認識システムにおける処理方式を、振込通知・振込照会・残高照会の各業務を例にして述べよう。図-3に示した各業務の処理の流れに従って、処理方式例を説明する。

(1) 振込通知：通知業務の処理は、センタからの通知データの発信により起動される。

① 中継用電算機 (RC) はセンタからの通知データを受信し、通知ファイルに格納する。

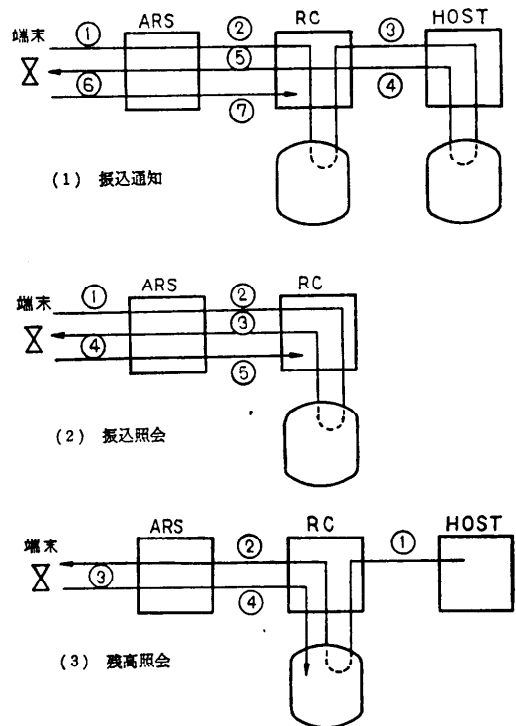


図-3 業務処理の流れ

② RC は通知ファイルからデータを取出し、音声応答認識システム (ARS) に送出する。

③ ARS はデータ中の電話番号により顧客を自動呼出しする。顧客が応答すると、音声応答により誘導しながら音声認識により顧客の確認を行い、音声により通知内容を伝える。

④ 通知が完了すると、ARS は RC に完了報告を行い、RC はそのデータの完了処理を行う。なお、顧客が不在などの理由で通知できなかった場合には、ARS はその旨 RC に伝え、RC はそのデータが未完である表示をし、後刻通知することとする。

(2) 振込照会：照会業務の処理は顧客からシステムへの着信により起動される。

① 顧客からの着信を検出した後、ARS は音声応答により顧客を誘導して、照会に必要なサービス種別や加入者番号などのデータを発声させ、音声認識により受信する。

② ARS は照会要求を RC に送信する。

③ RC は加入者番号等のチェックを行った後、通知ファイルからデータを読み出し、ARS に送信する。

④ ARS は照会結果を顧客に音声で通知する。

⑤ ARS は RC に照会の完了を報告する。

(3) 残高照会：残高照会の場合には、センタの元帳ファイルを参照する必要がある。

① 顧客からの着信検出の後、顧客からのデータを受信する。

② ARS は照会要求を RC に送信する。

③ RC は受信データのチェックの後、センタに照会要求を送信する。

④ センタでは加入者番号等のチェックの後、元帳ファイルを参照して、RC に照会結果を送信する。

⑤ RC は ARS に照会通知を指令する。

⑥ ARS は音声により照会結果を出力する。

⑦ ARS は RC に照会の完了を報告する。

(4) 音声認識装置に投入する情報：上述のような処理形式のサービスにおける顧客とシステムとの交話例を表-2 に示す。この種サービスで必要となる認識語は、

- i) 通知諾否の単語
 - ii) 加入者番号・暗証番号等の数字
 - iii) 通知繰返しの要否の単語
 - iv) サービス種別を識別する単語
 - v) 認識結果を確認する単語
- などである。サービス種別が多い場合には数字でコー

表-2 顧客とシステムの交話例 (振込通知)

顧客	システム
(応答)	(発信) 「こちらは××銀行です。振込の通知を致します。暗証番号をどうぞ。」
「××××」	「××××ですね。」
「はい」	「振込は×件です。 1番××繰から××円 ⋮ 以上合計××××円です。 これでよろしければおわりとおっしゃってください。」
「おわり」	「ありがとうございました。」

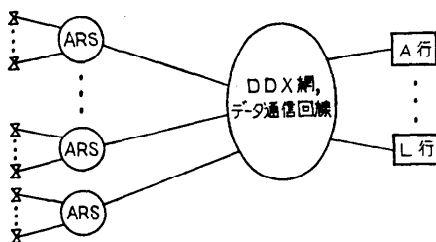


図-4 共同利用システムの将来構成

ド化することが可能であるので、バンキング・サービス用の認識語数は20語程度でよい。

4.3 共同利用システム

上述のようなシステムを各銀行が個別に設置してサービスを提供することは、設備投資に莫大な資金が必要となる、顧客の少ない地域にサービスの拡大をしにくい、顧客とセンタが離れている場合は通話料の負担が大きいの欠点がある。そこで、全国主要都市にこの種のシステムを設置し、多数の金融機関のオンライン・センタとデータ通信回線で接続する共同利用型音声照会通知サービスの計画が進められている⁷⁾。共同利用システムの構成案を図-4に示す。

共同利用システムは、全国レベルのサービスを経済的に実施できるほか、操作法が統一され顧客が利用しやすいなどの特徴を有しており、今後の発展が期待されている。

5. む す び

バンキング音声応答認識システムにおける音声認識の役割は、現状では情報入力のための押しボタン信号の代替手段にとどまっている。それでも上述のように、広く普及している回転ダイヤル電話機を用いて便利なサービスを提供できる。

音声という情報伝達媒体は一般の利用者にとって使いやすいが、発声における制約が多いとかえって使い

にくいものになる。この点では、現状の音声認識技術は十分なものではない。さらに使いやすいサービスを安価に提供するために、音声認識技術に対して次のような課題が与えられている。

(1) 認識性能の向上：電話入力音声の特性をさらに分析し、それに適合するよう認識法を改良し、認識性能を向上させる必要がある。

(2) 発声制約の緩和：連続単語の認識、問投詞等の余剰語の許容など、より使いやすい形式で入力できるよう認識機能を充実させねばならない。

(3) 小形・経済化：現在の所、音声認識装置は大きく、かつ高価である。今後の発展を期待するためにも、高速認識処理法の開発、LSI 等の新部品の適用などにより、装置の小形・経済化を進めることが重要である。

一方、銀行業務の効率化・サービス向上の点から見ると、音声応答認識システムの適用業務は今後拡大するものと予想される。本稿で述べた通知・照会関連の業務は、現在人手ではあるが電話で実施されている。従来窓口でのみ実施している振替依頼等の業務を電話から行う(在宅銀行利用)所まで広げると、一段とサー

ビス性が向上しよう。なお、この際個人識別のために話者認識技術が要求されるものと思われる。

参 考 文 献

- 1) 石井直樹：電話を用いた情報提供サービス，電子通信学会誌，Vol. 60，No. 10，pp. 1158-1163 (1977)。
- 2) 磯野智行：音声入力装置 DP-100，電子技術，Vol. 21，No. 12，pp. 16-20 (1979)。
- 3) 中津，長島：連続単語音声認識装置，日本音響学会音声研究会資料，S 79-54 (1979)。
- 4) 千葉，互理他：音声認識システム，大型プロジェクトパターン情報処理システム研究開発成果発表会論文集，pp. 157-165，日本産業技術振興協会 (1980)。
- 5) 千葉，互理他：特定音声認識の研究開発，大型プロジェクトパターン情報処理システム講演会論文集 (1977)。
- 6) 長島，中津：音韻単位の標準パターンを用いた実時間単語音声認識装置，日本音響学会音声研究会資料，S 78-22 (1978)。
- 7) 八橋主計：金融機関データ通信システムの技術的展望，データ通信，pp. 30-37 (1980-07)。

(昭和 55 年 12 月 18 日受付)