

第1章 序論

1.1 研究の背景と本研究の立場

電話による音声通信は人間同士のコミュニケーション手段であるだけでなく、音声対話サービスにおけるシステム/利用者間の入出力手段としても注目されている [1]。電話を用いることで利用場所と時間の制限を小さくでき、特別な道具やマニュアルも不要であることから、サービス性にすぐれたシステムの提供が期待できる。

音声対話システム(以下システム)と音声対話システムの利用者(以下ユーザ)との対話には、音声認識によりシステムにデータやコマンドを入力して所要のタスクを実行する場合と、音声メッセージを他のユーザに伝達するために蓄積する場合とが想定される。ここでは前者の対話を認識対話、後者を蓄積対話と呼ぶ。認識対話には銀行業務向け音声照会通知システム [2]、内線電話番号案内システム [3] など、蓄積対話には、ボイスメールシステム [4]、留守番電話システム [5] などの例がある。また、認識対話と蓄積対話の両方を用いるサービスも検討されている [6]。

認識対話の場合、音声認識の性能向上が主要な検討課題であった。音声認識の初期の課題は、発声ごとに単語時間長が非線形に異なることであったが、ダイナミックプログラミング (DP) の手法 [7] により解決された。不特定話者化、連続・大語彙化の問題は、HMM(Hidden Markov Model) など確率モデルを用いた手法 [8] により解決されつつある。電話回線経由の音声を認識するため、帯域制限や回線雑音を考慮した認識方式の検討も行われている [9][10]。

蓄積対話の場合は、当初メモリなどの音声蓄積媒体が高価であったため、無音区間の圧縮方法 [11]、高能率な音声符号化方式の研究 [12] が精力的に行われた。近年、フラッシュメモリなど半導体技術の進歩ともあいまって、テープレコーダに匹敵する経済性・録音時間と、ランダムアクセスによる迅速な再生制御の両立とが可能となった。

これらの技術的進歩により、部分的な制約はあるものの、認識対話、蓄積対話ともに性能および経済性の両面で、実用領域に到達したものと考えられている。

しかし、上記状況にも関わらず、音声対話サービスは日常生活の中に浸透しているとは言いがたい [5][13][14][15]。文献 [15] では音声認識対話システムが用いられない理由がアン

ケート調査を元に詳細に検討されているが、明確な理由は抽出されていない。蓄積対話については、文献 [5] で、システムの利用率の低さ (録音率 15%) が指摘されているが、その理由については明らかにされていない。技術面で実用領域に到達しているにも関わらず音声対話サービスが普及していないことから、従来の技術的課題以外の問題が存在していると考えられる。また、本来人間同士の会話手段である音声を、人間以外のものに対して用いること自体に抵抗を感じるユーザが存在する可能性もある。

上記状況を考慮し、本研究では、ユーザがシステムに対して発話することには、心理的な抵抗感など、ヒューマンインタフェース上の問題点が存在するという立場に立つ。本論文はそのような問題点を明らかにするとともに対話処理の高度化により改善し、音声対話サービスの普及促進に資するための検討結果を述べるものである。

1.2 従来のシステムの問題点

ここでは、ユーザがシステムに対して発話する際に生じる種々の問題点を整理する。システムの基本的な対話処理は図 1.1 の左部に示すような処理ブロックから構成される。はじめにシステムがガイダンスを送出する。以後、システムがユーザに対して送出手音声をガイダンスと呼ぶこととする。その後、通常はユーザが発話を開始するが、ユーザが発話を行わず、無言状態となることがある。その場合、システムはユーザの無言状態を検出し、ガイダンスを再度送出手音声を。ユーザが発話を開始した場合は、認識対話では音声の認識を、蓄積対話では蓄積を行う。認識または蓄積と並行して、ユーザの発話終了を監視する。認識対話の場合は、発話終了を検出後、認識結果の確認の対話を行う。

上記、①ガイダンスの送出手音声を、②無言状態検出手音声を、③認識または蓄積、④発話終了検出手音声を、⑤認識結果の確認、の処理ブロック毎にヒューマンインタフェース上の問題点を吟味する。

①ガイダンスの送出手音声をにおける問題点 ユーザに対し、システムは発話内容を指示するガイダンスを送出手音声を。従来のシステムでは、ガイダンスに含まれる情報は「〇〇をどうぞ」のように、必要最少限のみであることが多かった。システムとの対話に習熟したユーザの場合はこのようなガイダンスでも円滑な対話を行うことができる。しかし、不慣れなユーザや非協力的なユーザの場合は、無言状態となり電話を切断することがある。このような状況を改善するため、ユーザの心理的抵抗感を軽減し、発話を促進するガイダンスの検討等、対話シナリオの工夫が望まれる。

②無言状態検出手音声をにおける問題点 システムはユーザの発話の有無を監視し、一定時間以上発話が無ければ、無言状態と判断して、ガイダンスを再送する。従来のシステムで

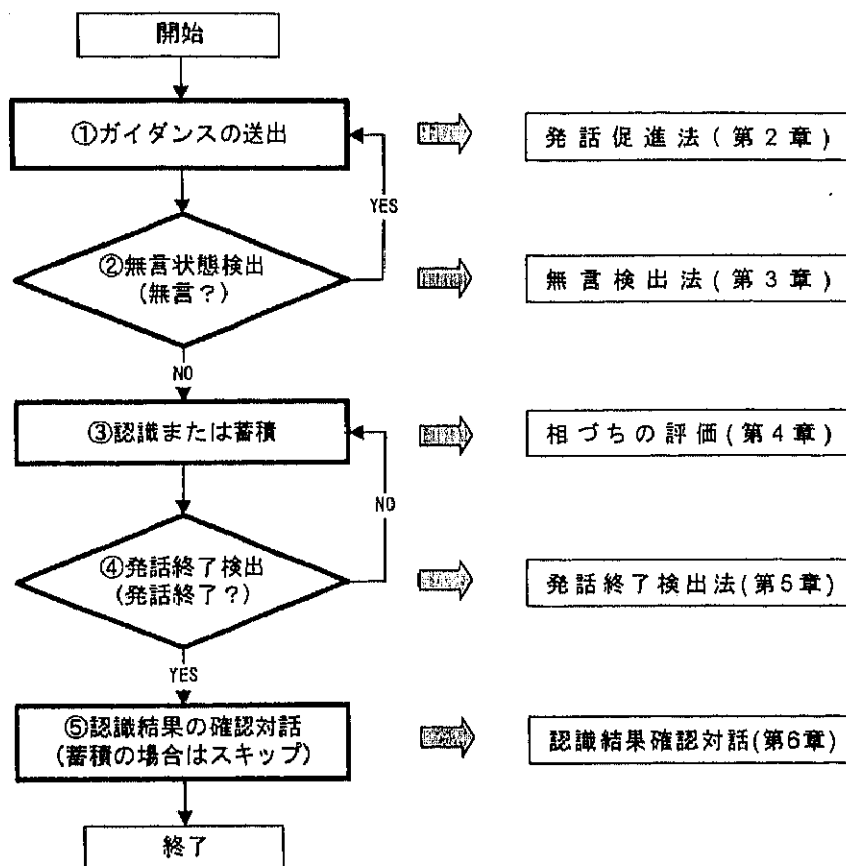


図 1.1: 基本的な対話処理項目と検討課題

は、無言を検出する時間長のしきい値として、経験的に得られた十分に長い一定値が設定されていた。その結果、システムが無言状態を検出する前に、ユーザが回線を切断する場合があります、対話完了率が低下する原因となっていた。ユーザが切断する前にガイダンスの再送を行うようにするため、できるだけ短いしきい値で無言検出を行う必要がある。

③認識時、蓄積時における問題点 ユーザがシステムに対して発話を開始すると、システムはユーザの音声認識または蓄積する。いずれの場合も、従来はシステムに対して協力的に発話を行うユーザのみを想定しており、システムに対して発話することに心理的な抵抗を感じるユーザを考慮していなかった。このようなユーザにとって、発話終了までシステムの応答や反応が全くないことは、極めて話しづらい状況の原因になると考えられる。

④発話終了検出における問題点 発話交代のタイミングは対話の円滑な進行に影響を与え

る。従来は、ユーザの発話開始後に観測される無音区間に対し、経験的に得られた一定のしきい値を設定して発話終了を検出していた。その結果、発話交代時のタイミングに間延び感が生じるという問題があった。また、しきい値を短くすると、発話終了前にシステムが発話を開始する両発話が生じるという問題があった。したがって、対話の状況に応じて発話終了検出のしきい値を制御し、正確かつ速やかに発話終了を検出する方法が必要とされる。

上記①～④は認識対話、蓄積対話の両方に関わる課題である。一方、認識対話のみに関わる課題として下記⑤が考えられる。

- ⑤認識結果の確認対話における問題点 音声認識を用いる対話システムでは、認識結果の正否をユーザに確認する対話が必要である。このためユーザにとって本来のデータ入力のための発話とは別に、正否確認のための発話が必要となり、発話回数が増加する原因となっていたことから、正否確認対話による発話回数の増加をできるだけ低減する対話手順の検討が必要である。従来のシステムでは、単純に認識尤度(あるいは距離値)の上位候補から正否を確認するなど画一的な手順が用いられており、尤度の大きさにより対話手順を制御するなど、不要な発話をできるだけ減らすという立場からみた検討が不足していた。

1.3 本研究の目的と検討課題

本研究は上記①～⑤の問題点を考慮し、以下に示す検討課題(図 1.1 の右部)に取り組むことで心理的抵抗感等の問題点を改善し、音声対話サービスの利用促進を図ることを目的とする。

- ①発話促進法の検討 システムを利用することなく切断するユーザの心理状態を検討する。その結果をふまえ、切断を防止し発話意欲を促進する方策を検討するとともにガイドランスの指針を示す。さらに主観評価実験によりその方策の有効性を検証する。
- ②無言検出法の検討 無言状態検出のしきい値を、発話前の無音区間長の実測値統計から求める手法を検討する。従来のシステムでは経験的に設定していたしきい値を、統計的な根拠を明確にすることで、より短い値に設定してタイミングよくガイドランスを再送することを可能とし、対話完了率の向上を図る。
- ③相づちによる対話の好ましさの検討 発話中の心理的抵抗感を軽減するため、相づちを用いた対話を検討する。はじめに人間同士の対話における相づちの発話タイミング

を測定する。その結果をふまえ、システム／ユーザ間の対話における相づちが、対話の好ましさに与える影響を主観評価実験により明らかにする。

- ④発話終了検出法 対話の間延び感や両発話など、システム発話タイミングの問題点を改善してテンポよい対話進行を可能にすることを目的に、速やかで確実な発話終了検出法を検討する。無音区間の個数および長さや発話終了確率との関係を検討することで、無音区間長しきい値をアプリケーションに合わせ最適に設定する方法を示す。
- ⑤認識結果の確認対話 認識結果の正否確認対話における発話回数の増加が、ヒューマンインタフェース上の問題点となることを考慮し、発話回数を低減する確認対話処理手法の検討を行う。計算機シミュレーション(モンテカルロ法)によりその有効性を示す。

上記検討課題のうち①～④は認識対話、蓄積対話の両方に関わる課題であり、⑤は認識対話のみに関わる課題である。これらの検討により、システムに対する発話の抵抗感等、ヒューマンインタフェース上の問題点を改善する。

また、認識対話を用いた場合、音声認識率が発話の抵抗感など焦点となる問題点の評価結果に影響を及ぼす可能性がある。そのような状況を避けるため、①～④の検討では蓄積対話を用いて評価実験を行うこととする。得られた評価結果は、認識対話にそのまま適用できる保証はないが、システムに対する発話という点で共通のタスクであり、問題点の改善の指標は得られるものと考えられる。

1.4 論文の構成

本論文の構成を以下に示す。

第2章では、発話促進のための対話シナリオとして、メッセージを一度に発話させる従来型に対し、複数の短いメッセージに分けて発話させる対話録音型と、発話したユーザに種々の情報提供を行う情報提供型の対話シナリオを提案して検討する。主観評価実験により、提案した方式は、従来型に比べてユーザ発話促進効果があることを示す。

第3章では、ユーザが発話を放棄し無言となった場合に対処するため、速やかな無言状態検出法を検討する。ユーザ発話前の無音区間長のデータを収集し、その実測値統計から、無言状態と判断する無音区間長しきい値を求める。また、対話内容に応じて、そのしきい値を変える必要があることを示す。

第4章では、ユーザ発話中の心理的抵抗感の軽減のため、システムが相づちを返す対話制御方法について検討する。はじめに人間同士の対話における相づちの発話タイミングの

実験結果を示す。それをふまえ、相づちがある対話とない対話とを主観評価実験により比較し、適切なタイミングの相づちが対話の好ましさを向上させることを示す。

第5章では、対話の間延び感の解消と両発話の防止という相反する問題を解決し、対話の円滑な進行を実現するため、応答性と信頼性のよい発話終了検出法を検討する。はじめに発話終了検出の基本的考え方を述べ、無音区間長に一定のしきい値を設定する従来法の問題点を指摘する。次に、検出された無音区間が発話開始後何番目に観測されたものであるかによって、無音区間長しきい値を可変制御する手法を示す。無音区間の数と長さから求められる発話終了確率が、システム運用上要求される設計値に達するときの無音区間長をしきい値とする方法を検討する。その結果、従来法に比較して、より速やかな発話終了検出が可能となることを示す。くわえて、システム発話タイミングに関する主観評価実験を行い、本手法を用いた発話終了検出処理により、ユーザに好まれるタイミングで対話の進行制御が可能であることを示す。

第6章は、認識型対話システムにおける認識結果の確認対話の発話回数を低減する手法を検討する。確認対話を省略する手順／正否確認対話を行う手順／再発話を指示する手順の1つを、候補単語の逐次型正解確率に応じて選択する対話制御法を提案する。逐次型正解確率とは、尤度分布の単語間の違いを考慮して求めた正解確率から、正否確認の過程で不正解とされた単語の影響を除いた値である。本手法と、直接的に尤度をしきい値とする従来法の発話回数を比較するため、モンテカルロ法を用いたシミュレーション実験を行い、本手法の有効性を示す。

第7章では論文のまとめと今後の課題を述べる。