

効率的な音声対話誘導方式に関する一考案

大森 久美子[†] 水澤 紀子[†] 東田 正信[†]

NTT 情報流通プラットフォーム研究所[†] (株) 国際電気通信基礎研究所[†]
kumiko, mizusawa@isl.ntt.co.jp[†], higasida@ctr.atr.co.jp[†]

1. はじめに

近年、音声認識技術の発達により、これを用いた製品やサービスが普及し始めている[1].

音声認識を用いた入力インターフェースは、ユーザにとって習得が容易である、他の入力手段と比べ入力速度が速い、手や目を使う他の作業を妨害しないという利点を持つ。しかし、現状の音声入力インターフェースは“まわりくどい”、“聞き間違えが多い”、“時間がかかる”などの理由からユーザ満足が得られていない。様々なサービスのフロントエンドとして音声を活用するためには、短時間で効率良くユーザの要求を確定し、要求された情報の提供を行うことが重要である。

本稿では、現行の音声認識技術の認識性能や認識精度を考慮した上で、ユーザにストレスを与えずに、ユーザ要求の確定・情報提供を可能にするための音声対話誘導方式を提案する。また、本提案方式の有効性検証のために行った評価試験及び結果について報告する。

2. 現行対話誘導方式

音声認識は、一定時間内に認識処理可能な語彙数には限界がある、及び 100% の認識率が得られる保証がないという 2 つの課題を有している。このため、大語彙を対象とした音声入力インターフェースでは、現在、以下のような方策を講じている。

- (1) 階層化した属性によって大語彙を分類し、各階層ごとに認識処理を行うことで、対象語彙数の減少を図り、上位から下位層へ絞込みを繰り返すことで情報を確定する。
- (2) 各属性ごとにユーザへ正解が確認できるまで正誤確認を繰り返し、誤認識を確実に修正する。上位属性を確定することで、次の対象である下位層の対象語彙数を減少させる。

このため現行方式では、情報確定のために、属性の階層数分の [入力要求・正誤確認] ターンを必要とし、確定するまで正誤確認を繰り返して先へ進むことができない。これが、まわりくどい、聞き間違えが多い、時間がかかるというユーザ不満の原因と考

えられる。

3. 提案対話誘導方式

現行の問題点を解決するため、[入力要求・正誤確認] の繰り返しによらない情報確定のための新対話誘導方式を考案し、ユーザの負担軽減を図った。

3-1. 提案方式の概要

本考案は、ユーザが要求する情報の特質に着目し、“思い込み認識”と“聞いているふり認識”という 2 つの新認識概念を併用することを特徴とする。

- ① 思い込み認識…ユーザの発話頻度の偏りを利用して、ユーザ発話の集中する発話語を上位から一定数選択し、優先的な認識対象語とする認識方法。
- ② 聞いているふり認識…思い込み認識が外れた場合、ユーザにその事実を知らせぬよう、正誤確認を行わず関連する別属性を尋ねる方向へ自然に対話を誘導する。2 つの属性に関する情報を利用して、認識候補を絞込む認識方法。

“思い込み認識”により、多くの属性値を有する下位の属性からの入力が可能になり、質問回数の減少、まわりくどくない質問順序の構成が実現する。一方、“聞いているふり認識”では、属性ごとの正誤確認、確定を行わないことから、対話全体として正誤確認回数の減少、かつ自然な対話が実現する。

3-2. 認識機能のモデル化と提案方式の詳細

思い込み認識、聞いているふり認識の併用による新対話誘導方式実現のために、人間が他人の発話を聞き取る際の状態を考慮して、上位に出力された認識候補の認識尤度の相対的な関係 3 種類にモデル化した。

A) 認識結果 1 位候補の尤度 \geq 規定閾値 a

ユーザ発話は、思い込み対象語の中にあり、かつ認識結果 1 位候補 (認識結果は正しい) であると判断。

B) 認識結果上位 n 位までの尤度和 \geq 規定閾値 b

ユーザ発話は、思い込み対象語の中におそらくあり、認識結果上位 n 候補に絞込めたと判断。

C) その他の場合

ユーザは思い込み対象語以外を発話したと判断。

これをもとに各モデルごとに、“思い込み認識”、“聞いているふり認識”を併用した対話制御プロセスを定めた。

A) 認識結果は正しいという判断に対しては、第1位の候補を、提示、正誤確認を行う。

B) 聞いているふり認識を併用し、関連する別属性との統合により、絞込めた上位 n 個の候補の中から正解を絞込む。例えば、 $n=3$ とした場合、上位 3 候補の尤度差によってさらに分類する。尤度差が大きい場合、有力候補とみなして、正解である確証を得るために、その候補の持つ他の関連属性をシステム側から提示し、ユーザに確認する。尤度差が小さい場合は、上位 3 候補の関連属性に認識対象を絞り、ユーザに対して、3 候補が識別できる関連属性の入力を要求する。

C) 聞いているふり認識による別属性認識の処理時間と並行して、思い込み対象外の候補に対し認識を行う。取得した別属性候補と相互に実在性を確かめることにより、正解を絞込む。

聞いているふり認識の結果に対しても、上記 3 モデルの適用を繰り返し、ユーザ要求の確定を行う。

4. 評価試験及び結果

本提案方式の有効性を検証するために、日本全国の住所を市区町村 (4,100 件) まで (例; 東京都港区, 神奈川県横浜市磯子区, 等) 確定することをタスクとしたプロトタイプシステムを試作した[2]。比較のために、都道府県, 市区町村の純に各地を行う現行方式も試作した。

音声の入出力にはマイク, スピーカを利用し, 音声認識エンジンとして NTT サイバースペース研究所開発の VoiceREX(Ver.1.8.5-beta)[3]を使用した。

提案方式は、あらかじめ市区町村を思い込み対象語として選定した。聞いているふり認識においては、都道府県 (対象 47), 大字 (18 万), 公園, 区役所などの目標物 (6,400) を関連別属性として尋ね, 市区町村との関連性の有無 (例; 都道府県名と市区町村名の候補は上位, 下位の関係にあるかどうか) を確かめることで、確定対象市区町村の絞込みを行った。

4-1. 評価試験内容

被験者 10 名に対して現行, 提案の両プロトタイプシステムを用いた対話試験を行った。思い込み対象語に含まれる市区町村 10 件, 含まれない市区町村 10 件の計 20 市区町村を確定対象として選択し、

各被験者に対して 20 件×2 方式の計 40 件の対話試験を実施した。各対話における、確定までの所要時間, 被験者発話回数を測定し, 被験者には各方式に対して満足度を 5 段階で評価してもらった (1: 不満, 3: 普通, 5: 満足)。

4-2. 評価試験結果

評価試験集計結果を以下の表 1 に示す。

表 1: 住所確定対話計 200 件 (10 名×20 件) の集計結果

	現行方式 200件	提案方式	
		思い込み対象 100件	思い込み 対象外 100件
平均確定所要時間 (秒)	54.0	26.9	40.8
平均被験者発話数 (回)	5.04	2.47	3.29
平均正誤確認回数 (回)	2.59	1.20	1.28
平均 5 段階評価	2.38	4.38	

提案方式において、確定までの所要時間, 被験者発話回数を現行方式と比較すると、思い込み対象語に含まれる住所は約 1/2, 含まれない住所は 3/4 に減少した。平均被験者満足度も、2.38 と “やや不満” レベルから “やや満足” を上回る 4.38 へと大幅に向上した。これは、被験者に対する以下のアンケート結果の数値的な反映と考えられる。

(問い) 提案方式の良いところは? (括弧は人数)

- 確定したいものから入力できる(9)
- 確定までの時間が短い(5)
- いちいち確認されない(2)

実際のアクセス状況を考慮した場合、104 (番号番号検索サービス) では、首都圏の番号案内を要求するコールが大半である。実用上は、港区や横浜市のような市区町村にアクセスが集中すると考えられる。そこで評価試験において 8 割のアクセス頻度の偏りを仮定すると、確定所要時間, ユーザ発話回数ともに、約半分に削減できることが考察できる。

5. まとめ

音声認識技術の性能を踏まえた上でのユーザ満足を得るための対話誘導方式を提案し、その有効性を確認した。今後は、CTI への適用を想定し、電話帯域音声利用時など、認識精度が更に劣化する環境下でも対処可能な、より頑健な対話誘導方式確立を目指す予定である。

参考文献

- [1] 小特集: 「音声対話システムの実力と課題」, 日本音響学会誌, Vol. 36, No. 11, 1999.
- [2] Scneiderman, B. ユーザインタフェースの設計, 日経BP社, 1995.
- [3] <http://www.hil.ntt.co.jp/Speech/REX/index.html>