4P-4

誤認識頻発状況下で選択肢列挙を行う音声対話システムとその評価

松山 匡子 † 駒谷 和範 ‡ 武田 龍 † 尾形 哲也 † 奥乃 博 †

† 京都大学大学院 情報学研究科 知能情報学専攻 ‡ 名古屋大学大学院 工学研究科 電子情報システム専攻

1. はじめに

音声対話システムでは，誤認識が頻発する状況下においても, 頑健にユーザの意図解釈を行い対話を遂行することが求められる. 我々は, システムが選択肢を列挙し, ユーザに指定させる対話（列挙型の対話）において, 音声認識結果に加えてユーザの**発話タイミング**を利用したユーザの指示対象理解 (図 1) を行っている [1]. 列挙型の対話では, 音声認識結果に誤りが含まれていてもユーザの発話タイミングから頑健に指示対象が推定でき, さらにシステムが選択肢を明示するので未知語による誤認識を避けることが可能である. システムが適宜列挙型の対話にユーザを誘導することで, 誤認識が頻発する状況下でも頑健に対話を遂行することが期待できる.

本稿では, 列挙型の対話にユーザを誘導するレストラン検索音声対話システムについて報告する. 列挙型の対話に誘導する際の課題は, (1) 選択肢の列挙に移行する状況の判定, (2) 列挙内容の選定の 2 つである. 前者に対しては雑音比や対話履歴に基づき誤認識状況下であるかの判定を行い, 後者に対しては文法検証結果や対話履歴を用いた列挙内容の選定を行う.

2. 京都市レストラン検索音声対話システム

本研究では, 京都市のレストラン検索をタスクとした音声対話システムを実装している. 本システムでは, 図 2 に示すように, 検索条件入力のフェーズ I, レストラン検索結果出力のフェーズ II, レストランの詳細情報出力フェーズ III の 3 フェーズからなる. ユーザがレストランの検索条件を入力すると, システムは該当する複数のレストラン名を列挙する. さらに, ユーザが指定したレストランについて詳細情報を尋ねると, システムが回答する.

本システムは, ヘッドセットマイクロフォンを用いずにロボットと対話を行うような実環境を想定しているため, 非接話マイクロフォンを用いる. このとき, マイクロフォンにシステムとユーザの発話が混入するが, Semi-Blind ICA 手法 [2] により, ユーザ発話のみを分離する. 分離したユーザ発話の音声認識精度は, 実験環境の残響や雑音の影響を受け低くなる.

3. 選択肢の列挙における課題とアプローチ

本システムは, 図 2 の対話フェーズ I, III において対話が進まない状況で, 列挙型の対話にユーザを誘導する. 本章では, 選択肢の列挙へ移行する状況の判定, 列挙内容の選定の 2 つの課題に対するアプローチについて述べる.

3.1 選択肢の列挙への移行条件

ユーザ主導形式で対話が進まない場合, つまり誤認識による対話の破綻のおそれがある場合に, 列挙型の対話にユーザを誘導する. 本システムでは, ユーザ発話毎に誤認識状況下であるかの判定を行い, 列挙型の対話に誘導する. 表 1 に, 判定に用いる特徴量を示す.

Spoken Dialogue System Based on Enumeration Strategy to cope with Frequent Incorrect Recognition and Its Evaluation: *Kyoko Matsuyama (Kyoto Univ.), Kazunori Komatani (Nagoya Univ.), Ryu Takeda (Kyoto Univ.), Tetsuya Ogata (Kyoto Univ.), and Hiroshi G. Okuno (Kyoto Univ.)*

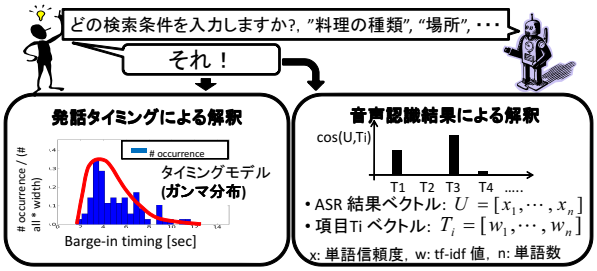

図 1: 発話タイミングを用いた解釈例

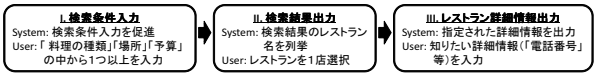

図 2: 対話フェーズ

V_{SNR} は, ユーザ発話毎の信号対雑音比 (SNR) 推定値 [3] である. ユーザ発話の音量が小さいことが誤認識の原因になる場合があるので, SNR 推定値を判定特徴量に用いる. 本システムでは, 予備実験から $D_{SNR} <= 20$ の場合, ユーザに「もう少し大きな声で話してください」と促し, それでも他の特徴の条件を満たす場合は, 選択肢の列挙に移る.

$C_{confirm}, C_{same}$ は, その対話のフェーズでは, 音声認識が困難な状況であるかを示す特徴量である. 例えば, フェーズ I で何度も確認発話が繰り返される状況であれば, 検索条件となる単語が未知語であるか, 音韻的類似度の高い単語が多く含まれ, 認識されにくい単語である可能性がある. また, 同じシステム発話が繰り返される状況, 例えば「検索条件を入力してください」が続く状況では, 対話が進んでいないと判定できる. 本システムでは, 表 1 の条件で, そのフェーズでは音声認識が困難であると判定し, 選択肢の列挙に移る.

C_{listup} は, その対話状況では音声認識が難しいかどうかの判定に用いる. それまでに一定回数以上列挙に移っていれば, そのユーザの声は認識されにくいか, 周辺環境が雑音の大きい状況にあるとみなせる. 本システムでは, 表 1 の条件が満たされる場合, 以降の対話はシステムが主導権を取り，列挙型の対話を行う.

3.2 列挙内容の選定条件

列挙内容は, 1) 検索条件 (フェーズ I) , 2) レストラン名 (フェーズ II) , 3) レストラン詳細情報 (フェーズ III) の 3 つに絞られる. 1) では, 料理の種類・場所・予算のいずれのスロットも埋まっていない場合は, スロット名を列挙し, まずユーザが条件を入力したいスロットを指定させる (図 1). また, 文法検証結果 [4] からどのスロットをユーザが埋めようとしているか判定できる場合は, その内容を列挙する. 例えば, ユーザが料理の種類を指定しようとしていると判定できれば, 料理の種類を列挙する. 料理の種類・場所・予算は選択肢が多いため, まず大カテゴリの内容を列挙し, さらに小カテゴリの内容を列挙する. 例えば, 料理の種類の大カテゴリに含まれる選択肢は,「和食」等であり, さらに「和食」の小カテゴリは

表 1: 誤認識状況の判定に用いる特徴

特徴	判定条件
1. V_{SNR}: SNR 推定値	$V_{SNR} <= 20$
2. $C_{confirm}$: 確認発話の回数	$C_{confirm} > 2$
3. C_{same}: システムの同一発話回数	$C_{same} > 2$
4. C_{listup}: 列挙への移行回数	$C_{listup} > 2$

表 2: 設定課題

	検索条件	詳細情報
1	祇園, 鶏料理（和食）	電話番号
2	金閣寺周辺, お好み焼き（和食）	閉店時間
3	京都駅周辺, イタリア料理 （洋食）	お店のウリ
4	八坂神社 の近く, 予算 3000 円	電話番号
5	京都大学 の近く, 創作料理（和食）	営業時間

「寿司」,「そば」等である. 2) のレストラン名は, 検索条件に合うレストランが複数あった場合, 五十音順に最初の 10 件を列挙する. 3) のお店の詳細情報は, システムが保持するお店の情報 (「お店の PR ポイント」,「電話番号」等 6 項目) を列挙すればよいため, 選定条件は特に設定しない.

4. 評価実験

表 2 にある 5 つの課題に沿って, (a) フェーズ II でのみ選択肢を列挙し, 音声認識結果のみで解釈するシステム, (b) 適宜選択肢を列挙し, 発話タイミングを併用して解釈する本システム, の順に対話データを収集した. 対話の制限時間は (a), (b) ともに 3 分とした. 被験者は, 指定された検索条件を満たすレストランを 1 つ選択し, そのレストランの指定された詳細情報を聞き出す. 表中の下線部は, システムにとって未知語である. これはシステムとユーザが保持する語彙に乖離がある場合に, システムが選択肢の列挙を行い, システムの持つ語彙をユーザに提示できることの効果の確認のために設定した. 被験者数は 20 代 ~60 代の一般男女 31 名, 発話総数は 6952 発話, 音声認識精度は 68.0%である‡.

4.1 タスク達成率

本システムで, より頑健に対話が遂行するかを確認するために, システム (a), (b) の各課題毎のタスク達成率 (タスクを達成した被験者数/全被験者数) を比較する. タスク達成は, 課題のレストランの詳細情報をそれぞれ制限時間内に聞き出せたかどうかで判断する. 各課題毎のタスクの達成率を表 3 に示す. 表 3 より, いずれの課題でも (a) に比べて (b) でタスク達成率の向上がみられる.

(b) を後に使ったことから, タスク達成率の向上分には, システムへの慣れも含まれる. しかし, (a) で未知語が認識されないことでタスク達成できず, (b) でタスク達成できた対話の中で, 列挙型の対話に誘導されることで初めてタスク達成できた対話は 49 対話中 27 対話存在し, 慣れの影響だけでなく本システムによる貢献も確認された.

一方, (a) でタスク達成でき, (b) でタスク達成できなかった対話は 10 対話だった. そのうち 5 対話は, (b) での制限時間を 5 分とする場合には制限時間内にタスク達成可能であり, 列挙型の対話に誘導することで, タスク達成までの対話時間が増加する状況が確認された. また 10 対話中 1 対話は, 列挙型の対話中にも誤認識が頻発し,

‡20dB のピンクノイズを重畳した音響モデル, CIAIR コーパスとシステムの保持する検索条件などの単語から作成した統計的言語モデル（語彙サイズ: 8239）を用いた.

表 3: 課題ごとのタスク達成率 (%) 被験者 31 名

	課題 1	課題 2	課題 3	課題 4	課題 5
(a)	19.4	71.0	3.23	22.6	16.1
(b)	35.5	77.4	38.7	45.2	63.3

表 4: アンケート調査結果

評価	1	2	3	4	5
列挙状況は適切か	2	8	4	8	9
列挙内容は適切か	1	5	10	5	10

フェーズ I で列挙型の対話が何度も繰り返されたために, 制限時間を 5 分に拡張してもタスク達成ができなかった. これは, 誤認識により大カテゴリ中の選択肢が誤って同定され, ユーザの意図しない小カテゴリの列挙に誘導され, ユーザが選択肢の列挙からなかなか抜け出せなかったためである. これに対処するために, 列挙型の対話を抜け出す機能をつける必要がある.

4.2 アンケート調査

表 4 に「選択肢を列挙する状況は適切だったか」,「読み上げる候補の内容が適切だったか」の問いに対するアンケート結果を示す. アンケートに対する評価は「適切でない」を 1,「適切だ」を 5 とし, 5 段階の評価を行う. 以後便宜的に 1, 2 を低い評価, 4, 5 を高い評価とする.

選択肢を列挙する状況については, 31 名中 17 名は評価が高く, 10 名は評価が低い. 選択肢の内容については, 31 名中 15 名は評価が高く, 6 名は評価が低い. 評価が低い被験者の意見としては,「意図していないところで選択肢の列挙にはいり, もとに戻れない」,「何度も同じ内容を繰り返し, (自分の発話が) 聞き取れていない」等, 選択肢の列挙に対して, うまく応答できない被験者の意見が多かった. 評価が低い被験者は, 全課題で平均して 27.5 回列挙に移っており (評価が高い被験者の平均は 21.9 回, どちらでもない被験者の平均は 23.3 回), 列挙への移行回数が多いことも低い評価につながったと考えられる. 一方,「認識されない理由がわからず不安になることが多かったので, もっと早くシステムから候補を読み上げるようアプローチして欲しい」等, 選択肢の列挙に移ることには肯定的な意見もあった.

5. まとめと今後の課題

本稿では, 発話タイミングをユーザの解釈に用いた音声対話システムについて述べた. 本システムによるタスク達成率の向上（平均 25.6 ポイント）を確認し, アンケート評価による本システムへの主観的評価について述べた. 今後, タスク達成率とアンケート結果の相関や提案手法の改良点の調査を続ける.

謝辞: 本研究の一部は科研費の支援を受けた.

参考文献

[1] 松山他. バージイン許容音声対話システムにおけるユーザ発話の分析と指示対象同定への応用. 音声言語情報処理研究会, 2010-SLP82-21, 2010.

[2] Ryu Takeda *et al.* Barge-in-able Robot Audition Based on ICA and Missing Feature Theory under Semi-Blind Situation. In *Proc. IEEE/RSJ IROS*, pp. 1718–1723, 2008.

[3] Chanwoo Kim *et al.* Robust signal-to-noise ratio estimation based on waveform amplitude distribution analysis. In *Interspeech*, pp. 2598–2601, 2008.

[4] 福林他. 音声対話システムにおける動的ヘルプ生成を指向した WFST に基づく文法検証によるユーザ知識推定. 人工知能学会研究会資料, Vol. SIG-SLUD-A703-09, pp. 45–50, 2008.