Vol.2012-SLP-92 No.2
2012/7/19

質問応答データベースを用いた聞き返し発話の検出に関する検討

三宅 真司[1,a)]　廣井 富[2,b)]　伊藤 彰則[1,c)]

概要：
生活の支援を行う会話ロボットのためのマルチタスク型の音声対話システムとして，複数タスクを容易に扱うための記述スクリプトに関する検討と，類似するタスクが存在した場合に必要に応じてユーザに聞き返しを行うための発話の検出手法に関して検討を行った．
システムでは状態遷移を伴った一問一答型の対話システムをサブシステムとして用いていることから，数ターン程度の確認を伴った対話を記述スクリプトによって容易に構築可能である．また聞き返しが必要な発話を質問応答データベースを用いて識別することから，話したい内容の追加時に同一内容の用例が存在した場合であっても，聞き返しをシステムが検出して行うため用例を自由に追加することが可能である．聞き返し発話は平均して 90[%] 近い検出精度を得ることができた．

キーワード：音声対話システム，質問応答，会話ロボット，対話管理，曖昧な発話

Detection of Utterances that Need Clarification Using a Question and Answer Database

Shinji Miyake[1,a)]　Yutaka Hiroi[2,b)]　Akinori Ito[1,c)]

Abstract:
We have conducted research on a multi-task spoken dialogue system for communication robots supporting livelihood. We studied definition script for dialogue control and a method of detection of user speech input with a low level of information needed for the identification of task. The proposed dialogue system script has an architecture based on state transition using a example-reply database as subsystems , drastically reduced an effort for system development. We conducted a experiment for the detection of utterance required by repeat. As a result, we obtained near 90[%] accutacy.

Keywords: spoken dialogue system, question and answering, conversational robot, dialogue control, ambiguous utterance

1. はじめに

我々は人間が普段生活する家庭環境を想定した生活の支援を行う会話ロボットの構築を目指している．音声は人間にとって身近なツールであり，人間同士が会話で物事を理解する事と同様に，ロボットと会話でやりとりをする事は有用であるといえる [1]．近年の音声認識技術の発展に伴い，観光案内など多くの音声対話システムが提案されている [2], [3] ことに加え，エージェントやロボットを用いたユーザフレンドリな対話システムの研究も盛んに行われている [4]．しかし音声対話システムは一般に，対話のドメインやタスクに特化して構築する場合が多く，既存の対話

[1] 東北大学大学院 工学研究科
Graduate School of Engineering, Tohoku University
[2] 大阪工業大学 工学部 ロボット工学科
Department of Mechanical Enginnering, Osaka Institute of Technology
[a)] miyake@spcom.ecei.tohoku.ac.jp
[b)] hiroi@med.oit.ac.jp
[c)] aito@spcom.ecei.tohoku.ac.jp

システムに対してタスクの追加や変更が困難であるという問題が依然として存在している [5]．この問題に対する解決策の 1 つとして，特定タスクのみを扱える対話システム（サブシステム）を複数用いたマルチタスク型音声対話システムが提案されている [6], [7]．つまりタスクの追加や変更は，個々の対話システムを変更するだけで行えるという利点があるといえる．

生活の支援を行うロボットのための音声対話システムでは会話ロボットの制御も複雑 [8] なため，サブシステムとなる対話システムを容易に構築できることに加え，対話制御を簡易に行えることも重要である [9]．また複数のタスクを扱うためマルチタスク型の音声対話システムであることも必要な条件である．我々は以前サブシステムの構築方法として，質問応答型の対話システムに対話状態を導入し，生活の支援を行うロボットのための音声対話システムを提案した [10]．本稿では，そのような質問応答型のサブシステムを用いたマルチタスク型の音声対話システムに関して検討を行った．

マルチタスク型の音声対話システムには，ユーザの最初の発話をどのサブシステムで処理するかを識別するという課題がある．サブシステムは互いに独立であるが，類似したサブシステムの場合にはタスク識別が誤ってしまう可能性が考えられる．ユーザ発話の曖昧性を検出し曖昧な箇所を推定し聞き返す枠組みは検討されている [11] が，生活支援を行う環境のようにサブシステムが類似する場合にタスク識別を行うための情報が十分かを検出するという検討は行われていない．想定している対話例を以下に示す．

想定している対話例

U1：電源を付けて
S1：テレビですか？エアコンですか？
U2：テレビ
S2：テレビの電源を付けるね

類似するサブシステムがない場合は上記の S1 の応答と U2 の発話は必要ないが，類似するサブシステムがある場合はこれらの聞き返しの発話は必要であるといえる．このようなタスク識別を行う際にユーザ発話に必要な情報が不足している発話を「聞き返し発話」と定義する．本研究ではこのような聞き返し発話を検出し，必要に応じて対話システムが聞き返すことで，複数タスクを扱う際にも円滑に対話を進行できるようにすることを目指す．目的とするシステムは聞き返しが必要かどうかの識別を行った後にタスク識別を行い，該当する個々のサブシステムで対話の処理を行うシステムである．概要を**図 1** に示す．本研究では音声認識エンジンには Julius[12] を，音声合成エンジンには Aquestalk2[13] を用いた．

本稿では，マルチタスク型の音声対話システムの構築のために開発した記述スクリプトと，質問応答データベースを用いた聞き返し発話の検出に関して述べる．

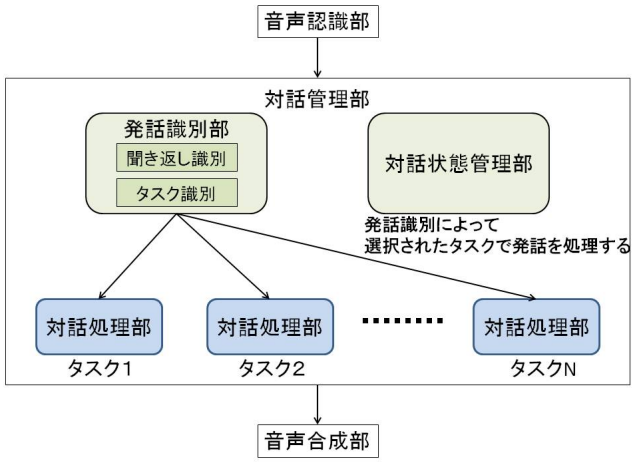


図 1 目的とするシステムの概要
Fig. 1 System Overview

2. 生活支援型の対話システム

2.1 生活支援型の対話

生活支援型の対話とは，例えば家電製品の操作や物を捨てたり持ってきたりするといったロボットに指示を行う形式の対話であり，家庭内で想定されるロボットとの会話を通して人間の支援を行うドメインの事である．このドメインでは，ユーザの発話に対して確認を行う数ターン程度のやりとりを伴った対話を想定していて，対話対象となる機器や物が複数存在していることや家電製品の操作など類似したタスクが多いこと，そしてタスクの追加や変更が多いことが特徴である．

これらの特徴から生活支援型の対話では，質問応答型の対話を用いることが有効であると考えられる．質問応答型とは一問一答型とも呼ばれ，ユーザの発話に対して予め決められた応答を質問応答データベースの中から検索して返答するタイプの対話のことを指す．数ターン程度のやりとりを伴った対話を記述する方法としては，フレームに基づくシステムと質問応答データベースに基づくシステムの 2 つが提案されている．フレームに基づいたシステムと比較すると質問応答データベースに基づくシステムは単純ではあるが動きが予測しやすく，音声対話研究の非専門家であっても対話内容の追加や変更などのメンテナンスが容易に行えるというメリットがある．

2.2 対話状態を伴った一問一答型の音声対話システム

一問一答型の音声対話システムは，ユーザの入力発話をシステムが予め保持した質問応答データベース中の発話用例文と比較し，最も一致した発話用例文に対応する応答候補文を音声合成により再生する [4]．発話用例文とは，**表 1**

表 1 発話用例文の例
Table 1 Sample of Example Text

タグ	発話用例文
#1001	テレビ+テレビ, を+オ/ヲ, 付け+ツケ, て+テ
#1001	電源+デンゲン, を+オ/ヲ, 付け+ツケ, て+テ
#1002	テレビ+テレビ, を+オ/ヲ, 消し+ケシ, て+テ

表 2 応答候補文の例
Table 2 Sample of Answer Text

タグ	応答候補文
#1001	テレビを付けるね
#1002	テレビを切るね

の様なユーザの入力発話を予測して書かれた用例文のことであり，応答候補文とは，**表 2** の様なシステムの応答文のことである．発話の多様性を確保するために，同一内容の複数の発話用例文と一つの応答候補文が対応している形となっている．

通常一問一答型の対話システムには対話状態が存在しないが，本研究では一問一答型の対話システムを複数繋げることによって対話状態を定めた [10]．認識結果から発話用例文を選択した時に発話意図が特定できるため，発話を単位として対話状態を定めることができ，タグを履歴として保持することで状態遷移先の記述を行うことができる．対話例を以下に示す．

対話例

状態 1
U1:#2003 ペットボトルを捨てて
S1:#2003 ペットボトルを捨てますか？
状態 2
U2:#3001 お願いします
S2:#2003 ペットボトルを捨ててきます

一問一答型の対話システムのそれぞれが対話状態に該当し，状態 2 では発話用例文として「お願いします」などを，応答候補文として「ペットボトルを捨ててきます」などを，状態遷移ルールは以下のような情報を保持している．

状態遷移ルールの例

#3001 履歴のタグを用いて応答，初期状態に戻る
#3002 現在のタグを用いて応答，1 つ前の状態に戻る

2.3 解決すべき課題

2.1 節で述べたように生活支援を行う環境では，対話対象となる機器や物が多く類似したタスクが多いことから，タスクの追加や変更を容易にする必要がある．

次に類似したタスクが多いことによる問題点として対象物が曖昧な発話が挙げられる．

対象物が曖昧なユーザ発話例

電源付けて，消して，上げて，下げて，変えて

上記のような発話例では，扱えるタスクがテレビのみの場合は処理するタスクを一意に決定することができるが，テレビとエアコンが扱えるシステムの場合，処理するタスクを一意に決定することができない．前者の場合は確認が不必要な発話であるといえるのに対し，後者の場合は確認が必要な発話であるといえる．つまりこれらの発話は扱えるサブシステムによって曖昧な発話かどうかが変化するため，類似するタスクが存在した場合には誤ったタスクに識別されてしまう恐れがあり，ユーザが意図しない操作を実行してしまう可能性が考えられる．したがって生活の支援を行う対話システムには以下の事が求められているといえる．

(1) タスクの追加と変更の容易性
(2) 必要に応じた確認発話

3. 質問応答型対話システムの構築

2.3 節で挙げた 1 つ目の問題点を解決するために，対話管理を行う設定ファイルの記述スクリプトに関して検討を行った．

3.1 対話管理の概要

設計した記述スクリプトは，色々なタスクのサブシステムを記述することが可能でありサブシステムの内部に対話状態を持つ．

従来のシステムで複雑となっていた意味フレームを用いることなく，対話状態のある対話システムを一問一答型に基づいて設計する．したがって対話状態を簡単に記述できるため，システムの動きが把握しやすくメンテナンスが簡易になるといえる．またそれらをサブシステムとして組み合わせることによって，タスクの追加や変更が容易な生活支援のための対話システムを容易に構築することができる．

利用者が構築するのは，発話用例データベースと応答候補データベースと状態遷移ルールの記述の 3 つであり，音声認識や音声合成に関して別途記述を行う必要はない．発話用例データベースは表 1 のような記述でまとめられていて，応答候補データベースは表 2 のような記述でまとめられている．状態遷移ルールには，表 1 の先頭に割り振られているタグと，2.2 節で述べた状態遷移先が記述されているものである．

3.2 記述スクリプトの構造木

図 2 に設計した記述スクリプトの構造木を示す．要素「block」はタスクを表し，要素「rec_state」は対話状態を表す．このような要素を組み合わせることによって，状態

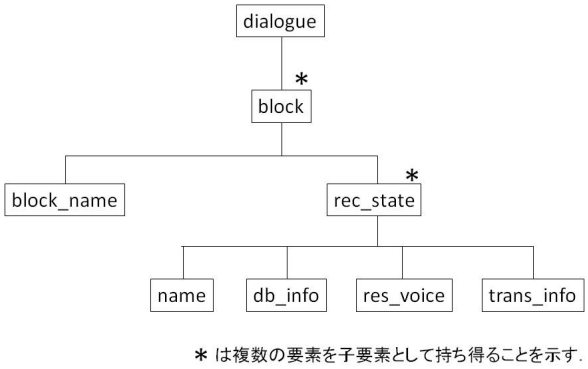


図 2 構造木
Fig. 2 Tree Structure

表 3 収集した対話データ
Table 3 Dialogue Data Collection

収集場所	防音室
教示	・テレビとエアコンが操作対象 ・発話例は各 5 例程度 ・言い回しなどは自由
被験者数	9 名
ユーザ毎の発話数	1 タスク当たり 20 発話を目安
総発話数	・テレビ：213 ・エアコン：226

遷移の伴った対話システムを構築することが可能である．

4. 実験データの収集とラベル付け

4.1 実験データの収集

質問応答型のシステムを用いて対話データの収集を行った．被験者は 9 名であり，類似する操作コマンドが多いと考えられるテレビとエアコンの操作に関する発話を収集した．収集した対話データの詳細を，**表 3** に示す．

4.2 ラベル付け

発話用例データベースと収集した対話データについて，聞き返しが必要な発話文に対するラベル付けを行った．発話用例データベースはテレビとエアコンとオーディオの操作に関連する 3 タスクと，物を捨てるタスクと物を指定した場所から探してくるタスクの合計 5 つのタスクで構成されている．ラベル付けの手順は以下の通りである．

- 各タスク間での同一用例文をマッチングによって算出
- 人手で付与

人手によるラベルの付与は，各タスク間で同一用例文は存在しないが意味的に同一な用例文に対するラベル付けである．ラベル付けは一人で行った．例えば「付けて」という用例文がテレビタスクには存在し，エアコンタスクには存在しない場合であっても，同一内容で別の言い回しの用

表 4 聞き返し発話の数
Table 4 Number of Utterances that Need Clarification

データの種類	通常の用例数	聞き返し用例数	総用例数
データベース	1318	47	1365
テレビ（open）	159	54	213
エアコン（open）	186	50	226

例文がエアコンタスクには存在するため，何を付ければいいのかがこの発話だけからでは分からないと判断し，聞き返しが必要とした．ラベル付けをした聞き返し発話の数を**表 4** に示す．

5. 識別のための特徴量

5.1 TF-IDF 法による重み

文書中の単語の重要度を測る手法として一般的な TF-IDF 法を利用して，ユーザ発話の重み付けを行う．文書 d に対する自立語 t の重みを $w(t,d)$ とすると，TF-IDF 法は式 (1) のように示される．

$$w(t,d) = tf(t,d)\log(\frac{N}{df(t)} + 0.1) \quad (1)$$

TF 項である $tf(t,d)$ はある文書 d における自立語 t の出現頻度を示している．IDF 項である $\log(\frac{N}{df(t)} + 0.1)$ は，自立語 t の文書出現確率の逆数に対数を取ったものであり，N は全文書数を，$df(t)$ は自立語 t が出現する文書数を示している．

ユーザ発話 s に含まれる自立語 t に対して求めた上記の値の和によって，評価値として文書スコアである $S_D(d)$ を式 (2) によって求める．

$$S_D(d) = \sum_{t \in C(s)} w(t,d) \quad (2)$$

$C(s)$ はユーザ発話 s に含まれる自立語形態素の集合を示している．また本研究では，1 タスクに関する発話用例データベースを 1 文書として扱っている．

5.2 不一致の自立語数

ロボットに指示をするような対話を想定しているため，自立語には動詞を含めている．そのため，動作を表す汎用的な自立語には高い重み $w(t,d)$ が付与されることから，ユーザ発話に含まれる自立語がどれほど文書スコア $S_D(d)$ に寄与しているかを判定するために，不一致の自立語数によるスコアリングを利用する．文書 d に対する不一致の自立語数を $c(t,d)$ とすると，不一致の自立語数によるスコアリングは式 (3) のように示される．

$$c(t,d) = \begin{cases} 1 & N(t,d) = 0 \\ 0 & otherwise \end{cases} \quad (3)$$

ここで，$N(t,d)$ は文書 d に含まれる自立語 t の数を示し

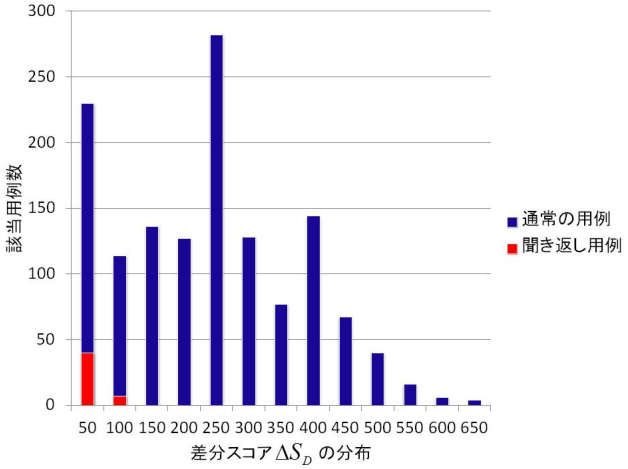


図 3 データベースの差分スコアの分布
Fig. 3 $S_D(d)$ Score Distribution of Database

表 5 自立語スコア $S_C(d)$ の分類
Table 5 Four Conditions of $S_C(d)$

		文書スコア $S_D(d)$	
	条件	1 位の文書	2 位の文書
自立語スコア $S_C(d)$	A	0	0
	B	0	1 以上
	C	1 以上	0
	D	1 以上	1 以上

ている．ユーザ発話 s に含まれる自立語 t に対して求めた上記の値の和によって，評価値として自立語スコアである $S_C(d)$ を式 (4) によって求める．

$$S_C(d) = \sum_{t \in C(s)} c(t,d) \tag{4}$$

5.3 特徴量の評価

5.3.1 文書スコアの分布

データベースと収集した対話データを用いて，文書スコア $S_D(d)$ を算出した．入力文の文書スコア $S_D(d)$ が最大のタスクを 1 位の文書として S_D^1 と表し，2 番目に大きいタスクを 2 位の文書として S_D^2 と表すとき，1 位の文書スコア S_D^1 と 2 位の文書スコア S_D^2 の差分を ΔS_D として，式 (5) を定義する．

$$\Delta S_D = S_D^1 - S_D^2 \tag{5}$$

データベースと収集した対話データのそれぞれについて，通常の用例と聞き返し用例に関する差分スコア ΔS_D のヒストグラムを**図 3**，**図 4** に示す．

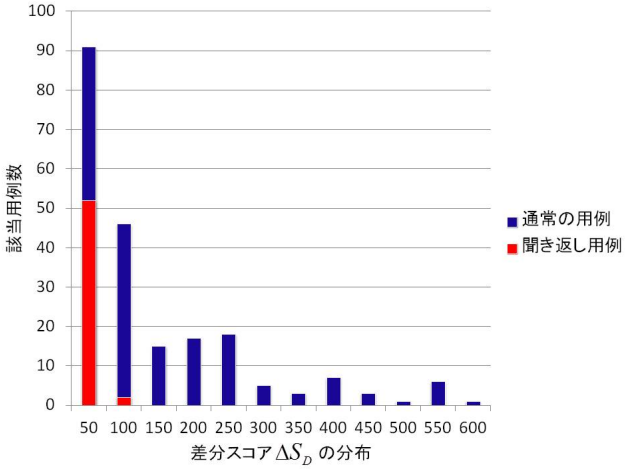
図 3，図 4 から，いずれの場合も聞き返しが必要となる用例の多くは，差分スコア ΔS_D が 100 以下の場合に多く分布していることが分かる．

5.3.2 自立語の分布

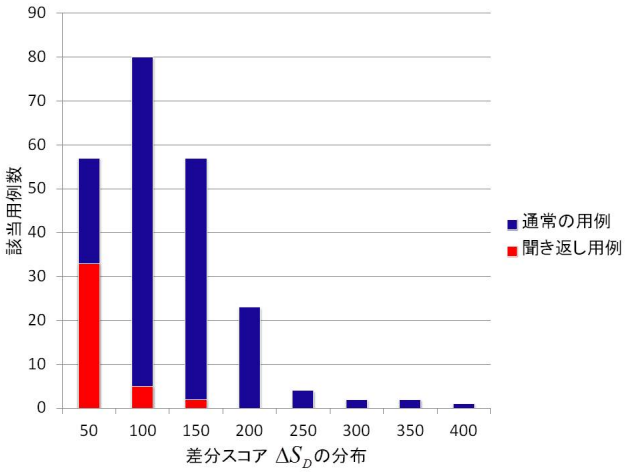
自立語スコア $S_C(d)$ を**表 5** のように 4 つに分類する．

データベースと収集した対話データのそれぞれについて，自立語スコア $S_C(d)$ に関して，通常用例と聞き返し用例がそれぞれの分類に含まれている割合を**表 6** に示す．

表 6 から，聞き返しが必要となる用例文は条件 A つまり，$S_C^1 = 0$ かつ $S_C^2 = 0$ の場合に多く含まれて，通常の用例は条件 B つまり，$S_C^1 = 0$ かつ $S_C^2 = 1$ の場合に多く含まれていることが分かる．



(a) テレビタスク
(a)Television Task



(b) エアコンタスク
(b)Air Conditioner Task

図 4 収集した対話データの差分スコアの分布
Fig. 4 $S_D(d)$ Score Distribution of Evaluation Data

6. 聞き返し発話の検出実験

2.3 節で挙げた 2 つ目の問題点を解決するために，聞き返し発話の検出に関して検討を行った．

6.1 検出アルゴリズム

聞き返しが必要かどうかを識別するためには，対話システムが入力発話に対して類似するタスクを最低 1 つ保持している場合を識別することが必要である．このため文書スコア $S_D(d)$ の値が近ければ，類似しているタスクを識別することが可能であると考えられる．一方で，汎用的な自立

表 6 自立語スコア $S_C(d)$ 分類時の通常用例と聞き返し用例の割合
Table 6 Ratio of Utterances that do/do not Need Clarification

自立語スコア	発話用例データベース		テレビタスク		エアコンタスク	
$S_C(d)$ の条件	通常用例 [%]	聞き返し用例 [%]	通常用例 [%]	聞き返し用例 [%]	通常用例 [%]	聞き返し用例 [%]
A	38.3	**61.7**	22.9	**77.1**	14.6	**85.4**
B	**99.2**	0.8	**84.6**	15.4	**97.2**	2.8
C	0	0	100.0	0	0	0
D	0	0	0	0	100.0	0

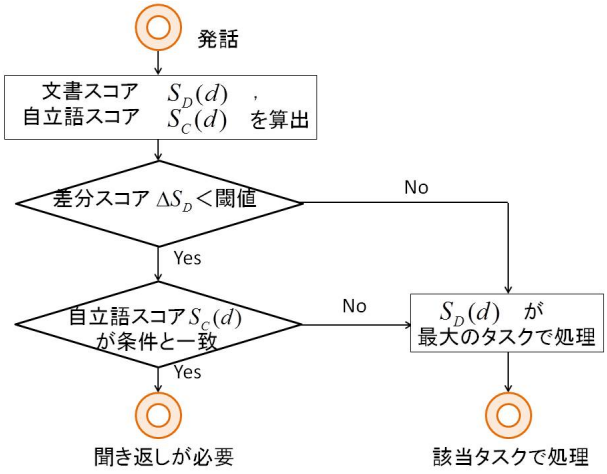


図 5 検出過程
Fig. 5 Method of Utterance Discrimination

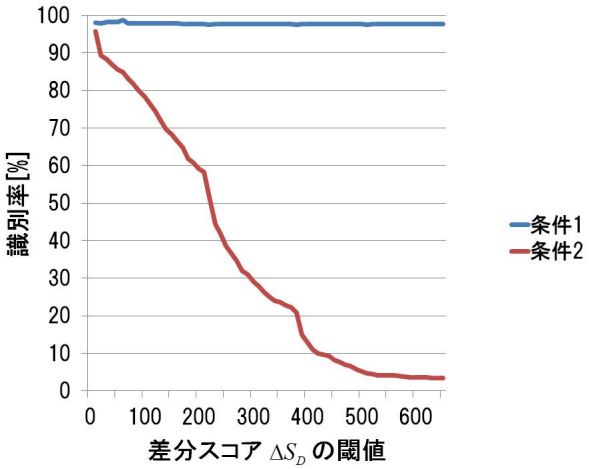


図 6 データベースの識別率
Fig. 6 Accuracy of Discrimination on the Database

語が含まれる場合は $S_D(d)$ の値が大きくなるため，自立語スコア $S_C(d)$ に条件を設けることで聞き返し発話かどうかを判定する．聞き返し発話の検出アルゴリズムを**図 5** に示す．

まずユーザの最初の入力発話を聞き返しが必要な発話か必要でない発話かの分類を行ために，差分スコア ΔS_D が閾値以下のユーザ発話を聞き返し発話の候補と判定する．次に 1 位の S_C^1 と 2 位の S_C^2 が，表 5 の条件 A に一致する場合の発話を聞き返し発話と判定する．

聞き返しが必要でないと判定された場合には，式 (6) により $S_D(d)$ が最大のサブシステムで処理を行う．$C(D)$ は文書の集合を示している．

$$\hat{S_D} = \max_{d \in C(D)} S_D(d) \tag{6}$$

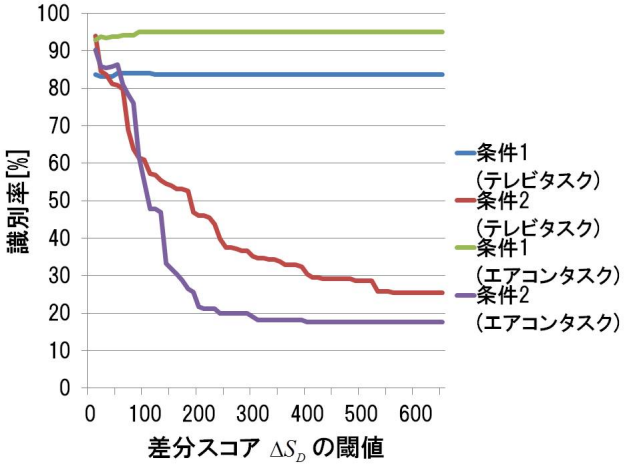


図 7 収集した対話データの識別率
Fig. 7 Accuracy on Discrimination on the Evaluation Data

6.2 実験条件

発話用例データベースと収集した対話データ (表 4 参照) を用いて，聞き返し発話の識別精度を調べる実験を，以下の 2 通りの場合に関して行った．

(1) 差分スコア ΔS_D と自立語スコア S_C^1 と S_C^2 を用いた場合
(2) 差分スコア ΔS_D のみを用いた場合

まず 5.3.1 項で示した差分スコア ΔS_D の閾値を 0 から 600 まで 10 刻みで変化させ，閾値以下のユーザ発話を聞き返し発話と判定した．次に 5.3.2 項で示した自立語スコア $S_C(d)$ は表 5 の条件 A，つまり $S_C^1 = 0$ かつ $S_C^2 = 0$ を満たす場合を聞き返し発話と判定した．評価には識別率を用いた．

6.3 実験結果

実験結果を**図 6** と**図 7** に示す．

条件 1 に関して，いずれの場合も閾値が 60〜100 の間の時に最も高い識別率となった．それぞれの最高値はデータ

ベースを用いた場合は 98.7[%]，収集した対話データを用いた場合はテレビタスクの場合に 84.0[%]，エアコンタスクの場合に 95.1[%] となった．条件 2 に関して，閾値を大きくすればするほど識別率が減少していく傾向がみられた．

これらの結果から，差分スコア ΔS_D が 100 程度の場合に聞き返し発話が多く含まれていて，また自立語スコア $S_C(d)$ を用いることによって，聞き返し発話の検出精度が高くなることが示された．

7. おわりに

生活の支援を行うロボットのための音声対話システムに関して，スクリプトを用いた対話管理による質問応答型のシステムの構築と，類似するタスクが存在した場合に聞き返しの識別を行う手法に関して述べた．収集したデータはタスク数が少ないものであったが，複数タスクを併用した場合において，高い識別率を得ることができたといえる．今後は，例えば家電製品の操作などはタスクとして似ているものが多いことから，データベースとのマッチング手法を更に検討することや，質問応答データベースの整備方法に関して検討を行っていく予定である．またスクリプトに関して，ロボットとの通信を伴った対話管理を実装するなどの改良も必要であると考えている．

参考文献

[1] 上田：ユビキタスホームにおけるサービスと対話インタフェースロボットの試作，電子情報通信学会モバイルマルチメディア通信研究会 (MoMuC), Vol. 105, No. 264, pp. 1–4 （2005）.
[2] 翠，河原，正司，美濃：質問応答・情報推薦機能を備えた音声による情報案内システム，情報処理学会論文誌, Vol. 48, No. 12, pp. 3602–3611（2007）.
[3] 駒谷，河原，清田，黒橋，P.Fung：柔軟な言語モデルとマッチングを用いた音声によるレストラン検索システム，情報処理学会研究報告, 2001-SLP-39-30, pp. 177–182（2001）.
[4] 西村，西原，鶴身，李，猿渡，鹿野：実環境研究プラットフォームとしての音声情報案内システムの運用，電子情報通信学会論文誌, Vol. J87-D-II, No. 3, pp. 789–798（2004）.
[5] T.Paek, R.Pieraccini：Automating spoken dialogue management design using machine learning:An industry perspective, Speech Communication 50, pp. 716–729（2008）.
[6] 長森，河口，松原，外山，稲垣：マルチドメイン音声対話システムの構築手法，情報処理学会研究報告, Vol. 54,pp. 45–51（2000）.
[7] S. Hahm, A. Ito, K. Awano, M. Ito and S. Makino: Utterance Classification for Combination of Multiple Simple Dialog Systems, Proc. Int. Symp. on Parallel and Distributed Processing with Applications Workshop, pp. 171–176 (2011)
[8] C.Shi，T.Kanda，M.Shimada，F.Yamaoka，H.Ishiguro and N.Hagita：Easy Development of Communication Behaviors in Social Robots，IEEE/RSJ International Conference on Intelligent Robots and Systems Taiwan, pp. 5302–5309（2010）.
[9] 成松，中野，船越，長谷川，辻野：ロボット・エージェント対話行動制御部構築ツール RIME-TK を用いた質問応答機能の実現，電子情報通信学会, SP2008-125,pp. 273–278（2008）.
[10] 三宅真司，廣井富，伊藤彰則：10 日間で作るロボット音声対話システム，ヒューマンインタフェースシンポジウム 2011，pp. 579–582 （2011）.
[11] C.Hori，T.Hori，H.Tsukada，H.Isozaki，Y.Sasaki and E.Maeda：Spoken Interactive ODQA System:SPIQA，In Proc.ACL, pp. 153–156（2003）.
[12] Julius：入手先 ⟨http://julius.sourceforge.jp/⟩.
[13] AQUEST：入手先 ⟨http://www.a-quest.com/⟩.