Vol.2009-SLP-77 No.15
2009/7/18

音声情報案内システムにおける質問応答データベース構築コスト削減の検討

竹内 翔大[†1] 川波 弘道[†1]
猿渡 洋[†1] 鹿野 清宏[†1]

実環境下で動作する音声情報案内システムに関して，質問応答データベース（QADB）を用いて応答を決定する，用例ベース応答生成手法が研究されている．この手法では，ユーザ発話を収集して質問応答データベース（QADB）を拡張することで，開発時に想定しないユーザ発話表現に対しても適切な応答が可能となる．しかしユーザ発話を収集して質問例を作成する作業は高コストである．本稿では用例ベース音声情報案内システムの QADB 拡張コストを削減するため，収集したユーザ発話からの未知質問例検出を検討した．提案する未知質問例検出手法は，どの既存質問例とも類似していない発話だけを検出する．収集発話データから未知質問例を選別することで，ラベル付けすべき発話データ量の削減と応答正解率の向上を両立できる．

Consideration of question and answer database construction cost reduction for speech-oriented information guidance system

Shota Takeuchi,[†1] Hiromichi Kawanami,[†1]
Hiroshi Saruwatari[†1] and Kiyohiro Shikano[†1]

In a real-environmental speech-oriented information guidance system, the example-based response generation method, which determinates a system response by using a question and answer database (QADB), has been investigated. When the QADB is expanded by collected user utterances, this system can respond the user utterance expressions which are not expected at the system design phase. Collection of example question from the user speech, however, takes much working cost. In this paper, we consider an unknown example detection of the collected user speech in order to reduce the QADB expansion cost for the example-based information guidance system. The proposed method detects the utterances which are not similar to all the existing example questions. This method achieves reduction of the number of utterances required to be labeled with maintaining improvement of the system response accuracy as well as full labeling.

1. 背　　景

人間にとって使いやすいインターフェースを持つ情報案内システムとして，音声入力を用いる音声情報案内システムが研究開発されている[1)–5)]．筆者らは実環境下で頑健に動作し使いやすい音声情報案内システムの実現を目指し，システムを市民センターや駅構内で長期運用して自然音声の収集やシステム長期運用に関する研究[6)] などを行ってきた．

近年，音声入力によるカーナビゲーションシステムの販売・携帯電話による音声地図案内サービスの開始など，音声情報案内システムが普及しつつある．しかし現時点では広く普及しているとはいいがたい．この問題には音声認識の精度・システム応答の精度という課題だけではなく，音声案内システムの新規開発・拡張の困難さという課題も含まれている．そこで本研究ではシステム拡張コスト軽減と応答精度改善の両立を目標とする．

実環境下でシステムを運用すると，不特定のユーザがシステムを利用するだけではなく，ユーザが開発者の想定しない質問を行うことが多い．特にユーザは，開発者の想定外の表現で質問するだけではなく，想定外の応答内容や時事的な内容を質問することもある[7)]．したがって実環境下で頑健に動作する音声情報案内システムを開発するためには，不特定話者の自然音声の認識技術のほかに，多彩な質問表現に対処できる対話処理手法が必要となる．また音声情報案内システムを長期運用する場合，ユーザに利用され続けることが重要である．そのためには運用状況に応じて新しい応答内容を提示するとともに，応答の精度を継続して改善することも求められる．したがって長期運用に堪える音声情報案内システムには，対処できる質問表現や応答内容を容易に拡張できる事が求められる．

システム開発・拡張の幾分容易な対話システムとして，用例ベース応答生成手法を用いるものが研究されている．この手法では，質問例とシステム応答のペア（QA ペア）のデータベースである質問応答データベース (QADB) を用い，ユーザ入力に最も類似した質問例を選択して適切な応答を決定する．用例ベース応答生成手法では QA ペアを作成して QADB に追加することで，応答内容やユーザ質問例の拡張を行える．また開発者の想定しない質問

†1 奈良先端科学技術大学院大学　情報科学研究科
Graduate School of Information Science, Nara Institute of Science and Technology

Vol.2009-SLP-77 No.15
2009/7/18

に対しては，運用中のシステムからユーザ発話および質問例を収集して QA ペアを作成することで対処できる．このように用例ベース応答生成手法では容易にかつ運用状況に応じたシステム拡張が行えるため，本研究では用例ベース応答生成を用いる音声情報案内システムを扱う．

運用状況に応じた音声情報案内システム拡張を行う際には，基本的には収集した音声をすべて人間が聴取し，ユーザ発話を選別して書き起こし情報や正解応答などのラベル付与を行う必要がある．しかしラベル付け作業は作業時間のかかる高コストな作業となる．このため運用状況に応じたシステム拡張を行うことは困難であることが多い．したがって本研究では，システム拡張コスト軽減のための手法を検討する．

ユーザ発話の選別に対しては，従来の雑音判別技術と年齢層判別技術で対処可能だと考えられる．また，質問例としてユーザ発話の音声認識結果を扱う QADB(ASR-QADB) を利用することで，書き起こし作業の省略や音声認識誤りの多少の対処を期待できることを報告した[8]．本稿では，既存 QADB に出現していない「未知」質問例を検出することで，新たにラベル付けすべきユーザ発話の件数を軽減することを検討した．

未知質問例検出は音声情報案内システムだけではなく用例ベース対話システム全般で適用可能な処理だと考えられる．また未知質問例検出の関連研究としてタスク外発話検出が挙げられるが，その違いを列挙する．

未知質問例検出は入力データが既存の質問例集合内に単に存在するかどうかを判定するもので，対話システムの外のシステムとして，QADB を拡張する際に実行される．学習データへの応答ラベルの付与は，判定時において必須ではない．一方タスク外発話検出では，入力データが応答可能かどうかを判定するもので，対話システムの構成要素の一部として，入力データに対してシステム応答を決定する際に実行される．タスク外発話判定モデルの学習データには応答ラベルの付与が必須となる．特に「応答不可能な既存の入力」に対して，タスク外発話検出では応答不可能として受理されることになるが，未知質問例検出では棄却される．未知質問例検出とタスク外発話検出の併用について，本稿では議論しない．

2. 音声情報案内システム「たけまるくん」

2.1 概　　要

本研究では「たけまるくん」[3] という音声情報案内システムを運用している．このシステムは生駒市北コミュニティセンターに 2002 年 11 月から設置され，以降現在まで 7 年近く運用を続けている．システムは GMM による雑音・音声の判別を行うことができ，雑音は棄却して応答処理を行わない．また大人音声・子供音声の不特定話者モデルを用いた並列デコーディングによって音声認識と年齢層の判別を同時に行うことができる．

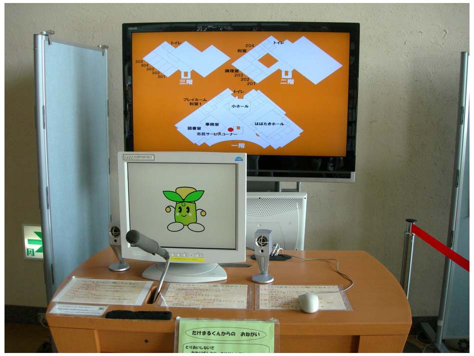

図 1　音声情報案内システム「たけまるくん」．
Fig. 1　Speech-oriented information guidance system “Takemaru-kun.”

対話戦略は一問一答で，大人 275 種または子供 285 種の応答からひとつを選択してユーザに提示する．応答情報は応答文の読み上げ音声のほか，地図やウェブページなどの画像情報とエージェントキャラクターのアニメーションが提示される（**図 1**）．応答内容はセンター施設案内および周辺施設・観光案内だが，ほかにも天気・ニュース情報などのウェブページの表示や，挨拶や「たけまるくん」自身のプロフィールなど，簡単な雑談も応答内容に含まれる．

「たけまるくん」では運用開始時から現在まで継続してユーザ自然音声を収集している．収集音声のうち最初の 2 年間の全発話について，人間の聴取により雑音性・年齢層・書き起こし・正解応答などの発話ラベルを付与し，音声データベースを整備している．

2.2 用例ベース応答生成手法

本節では「たけまるくん」で用いられる応答生成手法について述べる．たけまるくんは質問例と応答のペア（QA ペア）をデータベースとした質問応答データベース (QADB) を保持している．ユーザ音声を音声認識した後，有効音声に対して最近傍法によって QADB からユーザ入力に最も類似した質問例を選択し，選択された質問例に対応する応答をユーザへ提示する（**図 2**）．

音声認識時には入力音声の雑音判別および並列デコーディングによる大人音声・子供音声の大語彙連続音声認識が行われ，有効音声であれば音声認識結果と判別された年齢層が得ら

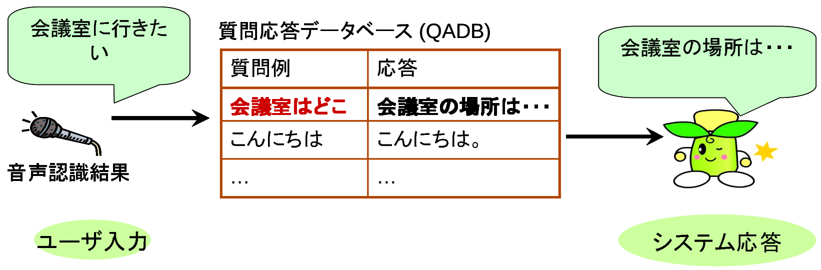

図 2　用例ベース応答生成.
Fig. 2　Example-based response generation.

れる．その後，年齢層に応じた QADB と入力音声認識結果とを比較し，類似度の最も高い質問例を選択し，選択された質問例に対応する応答をシステム応答として決定する．類似度の計算方法に関しては 3.2 節で述べる．なお従来の「たけまるくん」では，ユーザ発話の書き起こし文を QADB の質問例として利用していた．ユーザ発話表現と完全一致する質問例が存在しない場合でも，類似度の高い質問例を選択することで，ユーザ発話表現のゆれにもある程度対処できる．

2.3　システム拡張

用例ベース応答生成手法に基づく対話システムでは，QADB の質問例や応答内容を追加・更新することでユーザ入力に対する応答の精度を向上できる．特にユーザ発話から質問例を作成して QADB に追加することで，開発時には想定していないユーザ発話表現に対しても追加した質問例によって確実に応答できて，応答誤りを防げる．

収集されたユーザ発話から案内システム用の質問例を作成する場合，基本的には収集した音声をすべて聴取して書き起こし文や正解応答を付与し，QADB に追加するための QA ペアを作成する．ただし，雑音判別処理や年齢層判別処理を用いてユーザ発話から質問例として有効な発話を選別することも可能である．また質問例として音声認識結果を用いた QADB(ASR-QADB) を案内システム側で採用することで，書き起こし作業を省略するとともに応答の精度も少し向上させることができる[8]．

なお，一般的な用例ベース応答生成手法では案内内容を RDBMS で管理し，ユーザ入力に類似した質問例に対応するクエリを生成して応答内容を取得・応答生成する．しかし「たけまるくん」では処理を簡便にするため，応答文を質問例に直接対応づけ，QADB としてデータベース化している．

3.　提案するシステム運用法

3.1　運　用　系

本研究では音声情報案内システムの開発コスト軽減を目標とするため，本研究での提案は音声案内システムの運用方法に関するものとなる．本稿では「たけまるくん」のような用例ベース音声情報案内システムの運用方法について提案を行う．

ユーザ発話を収集して質問例を作成する場合，従来の手法ではユーザ発話をすべて聴取し，かつ書き起こし文と正解応答のラベルを付与する必要がある．しかし入力される発話の件数が多い場合は全発話へのラベル付与は膨大な時間を必要とするため，質問例の追加作業により応答の精度を継続して改善することは難しい．

ところで，QADB の質問例を拡張して応答の精度を改善するためには，未だ QADB に存在しない，未知の質問例を追加する必要がある．そこで，収集したユーザ音声から未知質問例を選別することで，ラベル付けすべきユーザ発話の件数を減少させることを提案する（**図 3**）．実際には「未知質問例を発声しているユーザ音声」を選別するが，便宜上「未知質問例」と省略して表現する．

提案する運用系をまとめると次のようになる．比較のため従来の運用系も述べる．

- 従来の運用
  QADB の質問例　収集したユーザ発話の書き起こし文
  ユーザ発話からの質問例作成　雑音判別，年齢層判別を併用
  付与する発話ラベル　書き起こし文，正解応答
- 提案する運用
  QADB の質問例　収集したユーザ発話の音声認識結果を用いる
  ユーザ発話からの質問例作成　雑音判別，年齢層判別，未知質問例検出を併用
  付与する発話ラベル　正解応答のみ

3.2　QADB 類似度スコアに基づく未知質問例検出

本稿では未知質問例検出手法として，ユーザ発話と既存質問例との最大類似度が閾値より低ければ（どの質問例にも類似していなければ）未知質問例だと判定する．最も単純な方法であるが，既存質問例に対して新たにラベルを付与する必要がなく，かつ現状の音声情報案内システムを変更することなく適用できる．

ユーザ発話 I と質問例 E との類似度 $s(I, E)$ は次の式により求められる[8]．

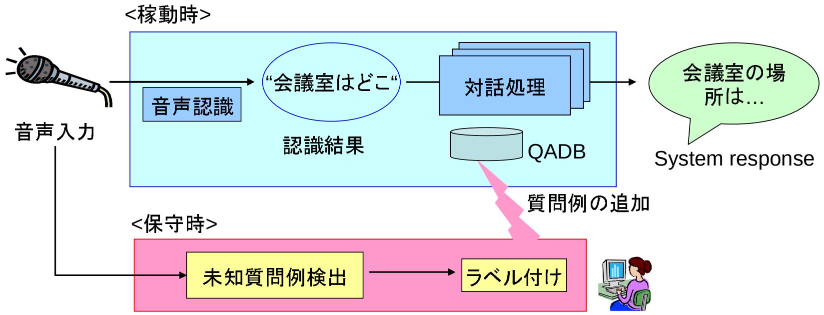

図 3 未知質問例検出を組み込んだ音声情報案内システムの運用.
Fig. 3 Operation of speech-oriented information guidance system with employing unknown example detection.

$w_I(k), w_E(k)$ 文章 I, E が持つ単語 k の単語重み（文章あたりの平均単語数）
$w(I, E)$ 共通する単語での平均単語数の総和
g_I I の一文あたりの単語数
Q 既存質問例の集合
$C(I)$ 発話 I の所属クラス（未知質問例か既存質問例）
θ 未知質問例判定のための類似度閾値

とするとき，文章 I, E の類似度 $s(I, E)$ を以下の式で与える．類似度の範囲は [0.0,1.0] を取り，文章 I, E が完全一致するとき値が 1.0 となる．ユーザ発話 I と既存質問例との最大類似度が設定した閾値 θ 以下となる場合，発話 I を未知質問例として検出する．

$$w(I, E) = \sum_{k \,\in\, I \cup E} \min(\ w_I(k),\ w_E(k)\ ) \tag{1}$$

$$s(I, E) \quad = w(I, E)\ /\ \max(g_I,\ g_E) \tag{2}$$

$$C(I) = \begin{cases} unknown & \text{if } max_{E \in Q}\ s(I, E) \leq \theta \\ known & \text{otherwise} \end{cases} \tag{3}$$

4. 実　　験

4.1 実験の概要

拡張データから未知質問例を検出して QADB 拡張に用いたとき，拡張データをすべて QADB 拡張に用いるときと同様にシステム応答の精度が向上するのかを検証した．具体的には，未知質問例検出によるシステム拡張コストと応答の精度との関係を評価した．

QADB 拡張を行う際には，応答正解率を優先する場合と拡張コストの軽減を優先する場合の両方が考えられる．前者に関して，拡張データをすべて QADB 拡張に用いるときと同様にシステム応答の精度が向上するのに必要なシステム拡張コストを求め，その低さによって未知質問例検出の有効性を議論することになる．後者に関しても，所望のコスト削減量において応答の精度が向上していれば未知質問例検出は有効だといえる．

応答の精度としては応答正解率を求め，システム拡張コストを表す尺度として，拡張データの総検出率を求めた．応答正解率は（正解応答の付与された）入力発話に対してシステムが正解できた発話の割合である．総検出率は，拡張データの総数に対する，未知質問例として検出された拡張データの件数の割合である．応答正解率は高いほど良い尺度で，総検出率は低いほど良い尺度だといえる．

4.2 実験手順

未知質問例検出の有効性を直接検証するため，雑音判別と年齢層判別は正しく行われたと仮定し，実験データには年齢層・書き起こし文・正解応答ラベルの付与された 2 年分の有効発話データを用いた．二区分の年齢層（大人・子供），二種類の入力データ形式（音声，書き起こし文）を組みあわせ，計 4 種の実験系を組んでいる．入力データ形式に関しては，入力と質問例に音声認識結果を利用する実験系と，入力と質問例に書き起こし入力を利用する系を組んだ．また各実験系ごとに，未知質問例検出の類似度閾値を変化させて総検出率と応答正解率の変動を求めた．なお，年齢層を区別するのは，使用しているモデルが大人・子供の二区分で構築されているためである．また入力データ形式の区分（音声，書き起こし文）に関して，前者では現実的な運用状況における提案法の有効性を，後者では音声認識誤りのない理想的な状況での提案法の有効性を検証する実験となる．

まず 2 年分の「たけまるくん」有効発話データを初期 QADB 作成データ・QADB 拡張データ・精度評価データに 3 分割した．初期 QADB 作成データは未知質問例検出による応答正解率の向上を確認するため，比較的少量である 1ヶ月分のデータを用いた．QADB 拡張データは 22ヶ月分のデータを用い，精度評価データは 1ヶ月分のデータを用いた．なお，正解応答は文献 8) 以降に修正を行ったものを用いている．**表 1** に実験データの件数を記載する．

QADB の作成では質問例と正解応答のペアを取り出し，重複を除いた．初期 QADB の作成後，QADB 拡張データから初期 QADB に対する未知質問例を選抜し，拡張データに

表 1 実験データ件数
Table 1 Number of data in experiments

初期 QADB 用データ	期間	2002 年 11 月	
	件数	大人	3,068
		子供	4,115
QADB 拡張用データ	期間	2002 年 12 月 - 2004 年 10 月 (2003 年 8 月は除く)	
	件数	大人	16,315
		子供	75,231
評価データ	期間	2003 年 8 月	
	データ	大人	1,053
	件数	子供	6,543
	応答の	大人	139
	異なり数	子供	206

表 2 未知質問例検出による QADB 拡張コスト削減と応答正解率の向上量
Table 2 QADB expansion cost reduction and response accuracy improvement by unknown example detection

	大人音声認識結果			大人書き起こし		
閾値	検出件数	総検出率	応答正解率	検出件数	総検出率	応答正解率
0.0	0	0.0	66.3%	0	0.0	68.9%
0.5	3106	0.190	73.5%	1983	0.122	73.1%
0.8	8312	0.509	78.5%	6364	0.390	81.1%
1.1	16315	1.0	78.7%	16315	1.0	82.0%
	子供音声認識結果			子供書き起こし		
閾値	検出件数	総検出率	応答正解率	検出件数	総検出率	応答正解率
0.0	0	0.0	53.1%	0	0.0	60.1%
0.5	18471	0.246	57.4%	8375	0.111	67.9%
0.8	50676	0.674	61.7%	36218	0.481	73.0%
1.1	75231	1.0	61.5%	75231	1.0	73.2%

正解応答を付与して拡張 QA ペアを作成した（実験では付与済みの正解応答を利用）．次に拡張 QA ペアを初期 QADB に追加し，初期 QADB を更新した．その後，精度評価データに対する初期 QADB の応答正解率を求め，同様に更新 QADB の応答正解率を求めた．未知質問例検出は閾値によって総検出率が変わるので，閾値を [0.0, 1.1] の範囲で 0.1 刻みで変化させ，閾値ごとに総検出率と応答正解率を求めた．全拡張データを確実に検出させるため，類似度の最大値 1.0 を超える値を閾値の最大値としている．

なお音声認識器には Julius[10] Ver 3.5.3 を用いた．音響モデルには JNAS モデルをたけまる音声データで再学習したモデル（3 状態 HMM, 2000 分布トライフォン PTM, 64 混合 GMM）を用いた．言語モデルにはたけまる音声データの書き起こしより作成したモデル（2gram + 逆向き 3gram，ウィッテン・ベル法でスムージング）を用いた．

4.3 結果と考察

実験結果を**図 4** に示す．横軸は拡張データの総検出率を表し，縦軸は評価データに対する応答正解率を表す．実線のプロットは音声認識結果の実験系，点線のプロットは書き起こし文の実験系による結果を表す．総検出率が 0.0 の時の応答正解率は初期 QADB の応答正解率を表し，総検出率が 1.0 の時の応答正解率は全拡張データをラベル付けして QADB 拡張した時の応答正解率を表す．グラフのプロットのそばに記載した数値は**表 2** に記載した応答正解率を示している．

図のグラフから，提案する未知質問例検出により，拡張データをすべて応答ラベル付けする場合での応答正解率を保ったまま，総検出率を 0.5 付近まで削減可能，つまりラベルづけする発話データ量を半減できることが分かった．また QADB 拡張コストの削減を優先する場合，総検出率を 0.2 程度に削減しても，応答正解率の向上は全付与の約半量を見込めることが分かった．

次に詳細な考察を述べる．入力・質問例に書き起こし文を用いる系では音声認識結果を用いる系よりも応答正解率が高く，また閾値を 1.0（完全一致）からわずかに下げても総検出率が大きく削減される．この結果は，書き起こし入力には音声認識結果で起こるような単語のゆれが生じないので，書き起こし入力が未知質問例である場合に類似度が大きく減少するためだと考えられる．また子供発話データの実験を大人での実験と比較すると，同じ類似度閾値では大人よりも検出量が幾分増加している．これは子供データの初期 QADB では評価データの発話表現をカバーできておらず，最大類似度が大人発話の実験よりも全体的に低くなったためだと考えられる．

以上の結果から，拡張データに対して未知質問例検出を行うことで，システム拡張コストの軽減と応答正解率の向上を両立させることができる．

5. ま と め

用例ベース音声情報案内システムのシステム拡張コストを削減するため，収集したユーザ発話からの未知質問例検出を検討した．この手法では収集したユーザ発話と既存質問例との類似度を計算し最大類似度が低い発話だけを検出する．収集発話データから未知質問例を選別することで，ラベル付けすべき発話データ量の削減と応答正解率の向上を両立できる．

本稿の実験では拡張データを一括して未知質問例検出にかけた．しかし実環境下の運用では，時事的な質問に対応するため逐次的に QADB を更新する必要がある．今後は逐次的に未知質問検出を行って QADB を拡張しても一括追加と同程度に応答正解率が向上することを確認したい．また本稿では雑音判別・年齢層判別が完全に行われた状況で実験を行っているが，今後は雑音判別・年齢層判別を含めた未知質問例検出処理の評価を行いたい．

参 考 文 献

1) A.L.Gorin, G. Riccardi, J.H.Wright, "How may I help you?," Speech Communication, vol.23, pp.113-127, 1997.
2) Victor Zue, Stephanie Seneff, James R. Glass, Joseph Polifroni, Christine Pao, Timothy J. Hazen and Lee Hetherington, "JUPITER: A Telephone-Based Conversational Interface for Weather Information," IEEE Transaction on Speech and Audio Processing, vol.8, no.1, pp.85-96, 2000.
3) Ryuichi Nisimura, Akinobu Lee, Hiroshi Saruwatari, Kiyohiro Shikano, "Public Speech-Oriented Guidance System with Adult and Child Discrimination Capability," Proc. of ICASSP2004, vol.1, pp.433-436, 2004.
4) Hiroya Murao, Nobuo Kawaguchi, Shigeki Matsubara, Yukiko Yamaguchi, Kazuya Takeda, Yasuyoshi Inagaki, "Example-based Spoken Dialog System with Online Example Augmentation," Proc. of ICSLP2004, Spec4402p-7, pp.3073-3076, 2004.
5) Cheongjae Lee, Sangkeun Jung, Seokhwan Kim, Gary Geunbae Lee, "Example-based dialog modeling for practical multi-domain dialog system," Speech Communication, vol.51, is.5, pp.466-484, 2009.
6) Tobias Cincarek, Izumi Shindo, Tomoki Toda, Hiroshi Saruwatari, Kiyohiro Shikano, "Development of Preschool Children Subsystem for ASR and QA in a Real-Environment Speech-oriented Guidance Task", in Proc. of EUROSPEECH 2007, pp. 1469-1472, 2007.
7) Jumpei Miyake, Shota Takeuchi, Hiromichi Kawanami, Hiroshi Saruwatari, Kiyohiro Shikano, "Language Model for the Web Search Task in a Spoken Dialog System for Children," in Proc. of WOCCI, Chania, Greece, October 2008.
8) Shota Takeuchi, Tobias Cincarek, Hiromichi Kawanami, Hiroshi Saruwatari, Kiyohiro Shikano, "Question and Answer Database Optimization Using Speech Recognition Results," INTERSPEECH 2008, pp.451-454, Sep, 2008.
9) http://sourceforge.jp/projects/chasen-legacy/ .
10) Akinobu Lee, Tatsuya Kawahara, Kiyohiro Shikano, "Julius — an Open Source Real-Time Large Vocabulary Recognition Engine," in Proc. EUROSPEECH 2001, pp.1691-1694, 2001.

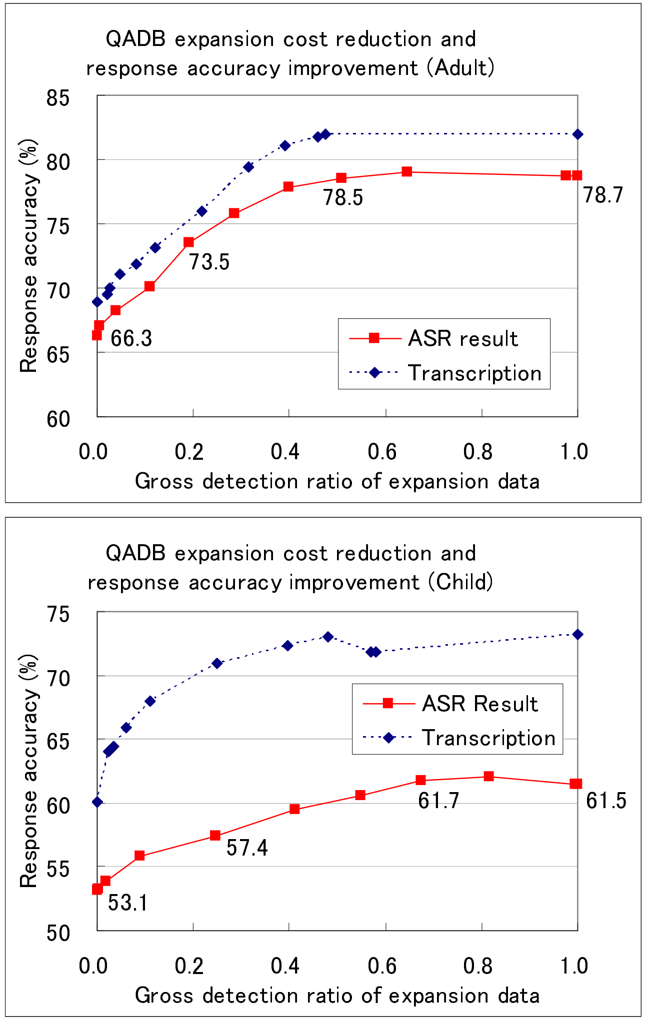

図 4 未知質問例検出による QADB 拡張コスト削減と応答正解率の向上量．
Fig. 4 QADB expansion cost reduction and response accuracy improvement by unknown example question detection.