

高精度な音声入力質問応答のための疑問表現補完

筒井 秀樹[†] 真鍋 俊彦[†] 福井 美佳[†]
酒井 哲也^{††} 藤井 寛子[†] 浦田 耕二[†]

音声入力を備えた質問応答検索システムでの誤認識に対するロバスト性向上のために、質問応答検索で重要となる疑問表現を補完する手法を提案する。我々がこれまで開発を行ってきた質問応答型マルチモーダルヘルプシステムに適用し、本手法の有効性を検証する。このシステムでは、ユーザの質問内容を理解し、ユーザが必要としている情報に対して的確に回答する質問応答技術を備えている。また、マルチモーダルコンテンツ検索技術を備え、映像・音声・取扱説明文などで構成される表現力豊かな回答を行うこともできる。これら2つの技術を融合することにより、より分かりやすい情報提供を実現している。たとえば、数量などを尋ねる Factoid 型の質問が入力されると、取扱説明書から質問応答検索により適切な回答が検索される。これに対し、方法などを尋ねる How 型の質問が入力されると、あらかじめ蓄積されたビデオデータベースから適切な説明ビデオを検索して表示することができる。しかし、最初のプロトタイプシステムでは音声認識誤りの影響により、検索処理がうまくいかず、適切な回答ができない場合があった。たとえば、具体的な時間や量をきく Factoid 型の質問は、音声認識誤りによって疑問表現の情報が失われて、Factoid 型の質問だと認識できなくなる。そこで我々は、音声認識誤りに対するロバスト性向上を目的とし、回答タイプが正しく判定されるように、音声認識結果を補完して検索する手法を開発した。本手法では、あらかじめ用意した疑問表現で音声認識結果を補完する。取扱説明書と音声認識結果との共起関係により候補を絞り込み、音声認識結果の発音の類似度を計算することにより補完する疑問表現を選択する。この手法により Factoid 型質問において平均逆数順位精度 (MRR) が 0.429 から 0.597 に向上した。

Recovering Interrogative Phrases for Effective Speech Input Question Answering

HIDEKI TSUTSUI,[†] TOSHIHIKO MANABE,[†] MIKA FUKUI,[†]
TETSUYA SAKAI,^{††} HIROKO FUJII[†] and KOJI URATA[†]

This paper proposes a novel method to improve retrieval accuracy in question answering with speech input by adding interrogative phrases to the search query based on co-occurrence of search terms in the query and interrogative phrases. We apply this method to a multi-modal help system and prove the effectiveness through an experiment. In our previous work, we developed a prototype of a speech-input help system for home appliances such as digital cameras and microwave ovens. Given a factoid question, the system performs textual question answering using the manuals as the knowledge source. Whereas, given a How question, it retrieves and plays a demonstration video. However, our first prototype suffered from speech recognition errors, especially when the Japanese interrogative phrases in factoid questions were misrecognized. To solve this problem the system supplements the speech query transcript with an interrogative phrase selected from a pre-determined list. The selection process first narrows down candidate phrases based on co-occurrences within the manual text, and then computes the similarity between each candidate and the query transcript in terms of pronunciation. Our method improves the Mean Reciprocal Rank of top three answers from 0.429 to 0.597 for factoid questions.

1. はじめに

近年、家電や AV 機器の高機能化やネットワーク化が進み、操作が複雑になってきている。また、多機能化が進み、ユーザがすべての機能を把握して使いこなすのが難しくなっている。製品には必要十分な取扱説明書が付属されているが、コールセンターへの実際の

[†] 株式会社東芝研究開発センター知識メディアラボラトリー
Knowledge Media Laboratory, Corporate Research & Development Center, TOSHIBA Corporation

^{††} 株式会社ニューズウォッチ自然言語処理研究室
Natural Language Processing Laboratory,
NEWSWATCH, INC.

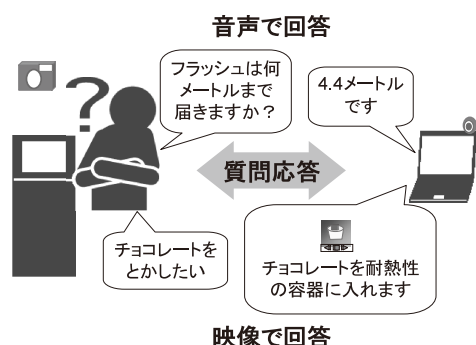


図 1 質問応答技術を備えたマルチモーダルヘルプシステム
Fig. 1 A multimodal help system based on question answering technology.

問合せ事例をみると取扱説明書に記述されている内容に関する問合せも少なくない。今後さらに取扱説明書のデータ量が増えるにつれ、知りたい情報を探せない、操作が複雑で取扱説明書を読んでも分からない、といった問題が増えると考えられる。

そこで、我々は、映像・音声・取扱説明文などで構成されるマルチモーダルナレッジによる表現力豊かなコンテンツの蓄積、検索技術と、質問内容を理解し、ユーザが必要としている情報に対して的確に回答する質問応答技術を融合し、より分かりやすい情報提供を可能とする質問応答型マルチモーダルヘルプシステムを開発し、プロトタイプシステムをタブレット PC 上で試作した²²⁾。

本システムには、知識源として家電の取扱説明書の説明文と、手段や方法を説明するためのアノテーション付き映像があらかじめ登録されている。我々は、ユーザがどの家電を使っているのかにかかわらず質問ができるシステムを目指しており、複数の家電の取扱説明書が区別なく本システムに登録されている。システムは入力された質問の種類によって回答の適切な形式/メディアを選択し、回答を提示する。たとえば、手段や方法に関する質問、すなわち、How 型の質問に対しては、映像や説明文で回答し、数量などを尋ねた Factoid 型の質問に対しては、音声（音声合成）と説明文で回答する（図 1）。

また、家電を操作中で手を離せないなど、ユーザの状況に依存しない検索システムを実現するため、本システムは音声による質問入力インタフェースを備えている。

しかし、ユーザからの音声による質問文に音声認識誤りが発生した場合、検索精度が著しく劣化してしまう。特に、Factoid 型の質問は、「何グラム」や「何メートル」など、質問の種類を判別するための疑問表

現が誤認識されると正しく質問の種類を分類できないという性質から、検索結果が音声認識誤りに大きな影響を受けることが分かった。

本研究では、音声認識誤りによる検索精度低下を最小限にとどめるため、質問文の音声認識結果を補完する手法について検討し、その効果について報告する。具体的には、音声認識誤りに大きな影響を受ける Factoid 型の質問において、質問の種類を判定するための疑問表現を補完する^{24),25)}。

以下では、2 章で関連研究について、3 章でマルチモーダルヘルプシステムのプロトタイプシステムの構成について、4 章で音声認識誤りの影響の分析について、5 章で提案手法について、6 章で評価実験について説明し、最後にまとめを行う。

2. 関連研究

近年、質問応答検索の技術は、TREC の QA Track²³⁾ や、QA@CLEF Track¹³⁾ や NTCIR の QAC Track⁵⁾ などで広く研究されている。Factoid 型質問文に対して 1 つの回答を答えるだけでなく、存在する複数の回答の組を答えるリスト型 QA 検索や、言葉の定義を尋ねる課題にも取り組んでいる。基本的な質問応答システムは最初に質問タイプ判定を行い、質問に対して何を答えるべきかを判定する。

一方、ユーザに対して自然なインタフェースを提供するため、情報検索の質問文を音声で入力するための研究も広く行われている。Barnett ら¹⁾ は音声による情報検索の比較実験を行っている。ここでは、問合せ内容を検索するシステムに対して音声入力インタフェースを備えるために、既存の音声認識システムが利用されている。TREC のトピックから選ばれた 35 件の質問文（101-135）をテストクエリとし、1 人の話者に音声で入力させた実験が行われている。また、Crestani²⁾ は、上の 35 の質問文を用い、音声入力による文書検索の精度向上のため適合性フィードバックを用いた手法を提案している。文書検索のモデルは音声認識とは独立しており、入出力を単純に接続したものであったが、音声認識誤りが検索精度に大きく影響することを指摘している。

検索文を音声で入力する場合、音声認識誤りの影響で検索精度が低下するため、その影響をできるだけ小さく抑えるために、さまざまな研究がなされている。藤井ら^{3),4)} は音声入力による文書検索を行っている。この中で、質問文中に未知語がある場合には、対象文書から語彙を抽出している。また、西崎ら¹⁶⁾⁻¹⁸⁾ は、音声入力されたキーワードによりニュース音声データ

ベースを検索するシステムを開発した．この中で，音声認識の次候補まで用いキーワード候補間の関連性を用いることで，音声認識誤りに対する耐性を向上させている．

また音声入力を適用した質問応答システムについても研究がなされている．堀ら^{6),7)}は新聞記事のような巨大なコーパスを利用した音声入力による質問応答システムを開発している．この中で，質問文を解析し，曖昧な点がある場合には，それをユーザに問い返すための対話文を生成する．

また製品の使い方をナビゲーションする取り組みとして，清田ら^{9),10)}は音声入力を備えた Microsoft Windows システムに対するオンラインヘルプシステムを開発した．このシステムは自然言語による質問応答が可能である．ユーザが入力した質問文に曖昧性がある場合，システムはその曖昧性を解消するための質問文を生成し，対話的に聞き返しを行うことによりユーザが求める答えにナビゲートする．音声認識誤りを解消するためにも対話が行われ，音声認識誤りの箇所をユーザに問い返す．検索対象から学習した単語 N-gram モデルによる単語パープレキシティを使用して検索整合度を計算し，検索に決定的な影響を与える語句を調べる．また，音声認識結果の N-best 候補を用いた場合の検索結果の違いから，N-best 候補間の相違箇所を提示してユーザに確認している．このシステムでは「～って何ですか」など用語の意味や定義を質問する What 型，「～の使い方」など手順や手段を質問する How 型，「～ができない」など問題や症状を述べ解決策を質問する Symptom 型の質問に着目している．

3. マルチモーダルヘルプシステムの構成

プロトタイプシステムの構成図を図 2 に示す．本システムはサーバ・クライアントモデルで構成され，クライアント PC でのヘルプインタフェース(1)と，サーバ側のヘルプマネージャ(2)，ビデオ検索エンジン(3)，質問応答エンジン(4)，メディアサーバ(5)からなる．システムを利用している様子を図 3 に示す．

クライアント上のヘルプインタフェース(1)は内部に音声認識エンジンを持ち，音声で入力された質問の音声認識を行う．また，音声合成を用いたり，検索されたビデオコンテンツの再生により，質問に対する回答の提示を行う．本システムではより自然な質問応答ができるようにするため，Factoid 型の質問と How 型の質問の両方を受け付ける．

音声認識エンジンには連続音声認識エンジンを用いる．音声認識エンジンにはほかにグラマベースの音

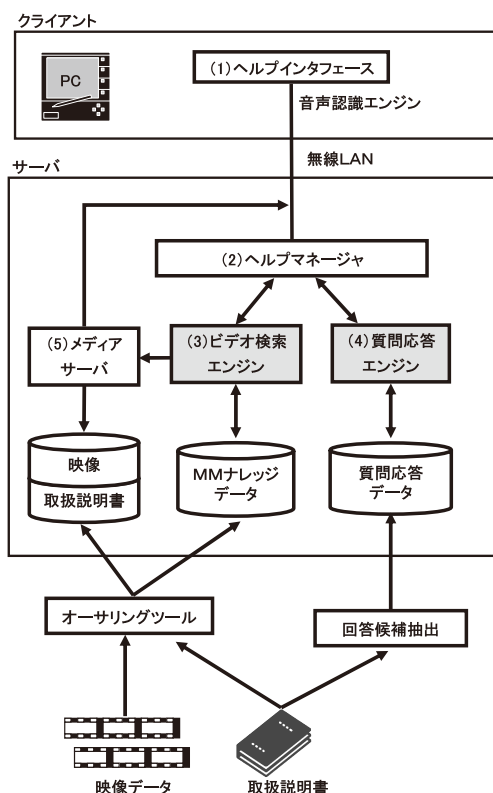


図 2 マルチモーダルヘルプシステムの全体構成
Fig. 2 Configuration of our prototype help system.



図 3 プロトタイプシステム全景
Fig. 3 The pilot model of a help system for home appliances.

声認識，キーワードスポッティングが考えられるが，グラマベースの音声認識の構築には認識辞書や文法の作成にかかるコストが高い²⁶⁾．またキーワードスポッティングでは，あらかじめ認識する語を決めておかなければならない．さまざまな質問に対応できるようにするため，本システムでは連続音声認識を採用した．音声による質問文入力と，音声合成による検索結

表 1 質問タイプ
Table 1 Question type.

質問タイプ	検索エンジン	質問内容の例
Factoid 型	質問応答エンジン	数字 (時間, 量など), 操作部の名前, 場所 例: スリープモードで電源が切れるまでの時間は何分
How 型	ビデオ検索エンジン	一連の操作方法 例: 夜景を撮りたいんだけど

表 2 Factoid 型疑問表現抽出ルール例

Table 2 Example of key expression extraction rule for factoid question.

(1)	何+ (キロ ミリ) ? + グラム
(2)	何+ (時 分 秒)

果の読み上げに, 東芝 LaLaVoice2001¹²⁾ の連続音声認識, 音声合成機能を利用した. LaLaVoice2001 は東芝が独自に開発した大語彙連続音声認識 (LVCSR: Large Vocabulary Continuous Speech Recognition) システムである. 本システムは音声の音響的特徴を判定する音響モデルと, 入力された文章の言語的妥当性を判定する言語モデルに基づいて, 入力音声を 1 字 1 句すべて認識結果として出力することができる. LaLaVoice2001 の言語モデルは N-gram ベースの統計的言語モデルであり, 1,000 万文超の新聞記事コーパスから抽出された bigram, trigram で構成される. また認識語彙数は約 10 万語である. 音響モデルは KL と GPD の競合学習¹⁹⁾ による SMQ/MHH¹⁴⁾ に基づいている. LaLaVoice2001 は話者の音声を学習する必要がない不特定話者認識システムである.

サーバ上のヘルプマネージャ (2) はヘルプインタフェースから受け取った質問文のタイプ判定を行い, この判定結果に従ってビデオ検索エンジン (3) と質問応答エンジン (4) のどちらかを呼び出す (表 1).

ヘルプマネージャは質問文からルールベースで疑問表現を抽出し, 数量 (時間, 分量など) または操作部の名前や場所を尋ねる疑問詞が含まれている場合は, その質問文を Factoid 型と判定する. Factoid 型疑問表現抽出ルールの例を表 2 に示す.

表 2 (1) には質問文中の「何グラム」「何キログラム」などの疑問表現が合致し, 表 2 (2) には質問文中の「何分」「何秒」などの疑問表現が合致する. たとえば質問文「スリープモードで電源が切れるまでの時間は何分」は, 表 2 (2) のルールに合致するため, ヘルプマネージャによって Factoid 型と判定される. ヘルプマネージャはすべての Factoid 型疑問表現抽出ル



図 4 映像の提示画面

Fig. 4 The screen for reproducing a video stream.

ルに適合しなかった質問文を How 型と判定する.

ビデオ検索エンジン (3) は, 複数の映像, 音声, 画像のファイル (リアルストリーム) から MPEG7¹⁵⁾ により必要部分をつなぎ合わせた仮想ストリームが検索対象として MM ナレレッジデータに登録されており, 時間軸上に関連づけられたテキスト情報 (アノテーション) を検索する. 検索にはベクトル空間モデルの文書検索をストリームデータ向けにカスタマイズした手法を採用している²¹⁾. How 型の質問文に対し, 検索結果の 3 位までがヘルプインタフェース上に提示される. 1 位の結果が映像であれば, メディアサーバ (5) を通して即座に再生画面が表示され (図 4), 取扱説明書のページ画像であれば, そのまま表示されるようにヘルプインタフェースの画面が構成される. 2 位と 3 位の回答候補は, ウィンドウの下部にそのタイトル情報のみが表示される²²⁾.

質問応答エンジン (4) には, 我々が独自に開発した質問応答システム ASKMi²⁰⁾ を利用した. ASKMi はあらかじめ取扱説明書から数量, 操作部の名前, 時間表現などの情報を抽出して質問応答データに登録しておく. これら回答候補のデータは抽出元のページ (回答の根拠文書) と関連づけておく^(8), 20). 質問に対する回答は信頼度の高い順に 3 位までの候補をヘルプインタフェース上に提示する¹¹⁾. 1 位の回答については音声合成による読み上げと, 根拠ページの提示が行われる (図 5).

4. 音声認識誤りの影響

前章のプロトタイプシステムを用いて, 音声認識誤りの影響を調べる評価実験を行った. この実験では 96 件の質問セットを構成し, 男女 10 人の被験者によ

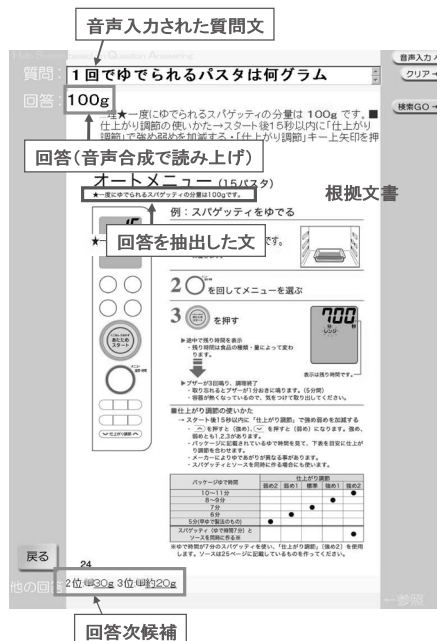


図 5 Factoid 型質問に対する回答画面

Fig. 5 Help interface outputs search results for Factoid question.

り読み上げた全 960 件の質問音声を入力する。

質問セットは、実際に家電のコールセンターへ問合せのあった質問をオペレータが記録した 1,657 件から、質問部分を人手で抽出して 1 文になるように整形し作成した。言い間違いや言い淀みは含まれない。また我々は全質問の中から Factoid 型と How 型の質問文に着目し、実際の問合せと同じ割合の Factoid 型 40 件, How 型 56 件で質問セットを構成した。質問文の平均文長は 14.4 文字である。

さらに、音声認識誤りによる検索精度劣化を見るために、後で述べる表 4 におけるキーボードによる質問文入力時の検索精度が 0.7 付近になるように質問セットを選択している。

検索対象としてオープンレンジとデジタルカメラの取扱説明書と利用方法の説明映像を用いた。これら 2 つのマニュアルは区別なく検索対象データベースに登録されており、ユーザはどちらの機器を使っているのかにかかわらず質問することができる。

音声認識システムの学習には、検索対象の取扱説明書を利用し、LaLaVoice2001 の「単語自動登録&学習機能」で登録した。この学習機能は文章を入力することにより未知の用語やいい回しについて自動的に学習することができる。学習の結果、取扱説明書から 415 語が未知語として学習された。式 (1) で計算した

表 3 音声認識誤りの例

Table 3 Example of questions and their misrecognized transcripts.

Factoid 型	正	動画を撮影するときの画像サイズは何ピクセル
	誤	動画を撮影するときの画像サイズ阿南ピクセル
Factoid 型	正	シャッター速度は何秒
	誤	シャッター速度は難病
How 型	正	予熱の設定はどうやるの
	誤	余熱の設定はどうやるの

表 4 音声認識誤りによる検索精度劣化

Table 4 Summarizes the results.

	(1)	(2)	(3)=(2)/(1)
Factoid 型質問	0.742	0.429	57.8%
How 型質問	0.732	0.625	85.4%

- (1) キーボード入力
- (2) 音声入力
- (3) 検索精度劣化の割合

実験環境における文字単位の平均音声認識精度は、学習前 0.788, 学習後 0.791 となった。また、質問文中の検索ターム単位の音声認識精度は、学習前 0.761, 学習後 0.793 となった。

$$1 - D_t / N_t \quad (1)$$

ここで、 D_t は認識結果と質問文の文字単位の編集距離であり、 N_t は質問文の文字数である。音声認識結果は漢字かな混じり文で出力され、検索に用いるためには表記が正しいことが必要であるため、音声認識精度の計測には文字単位の認識精度を用いた。

質問とその音声認識誤りの例を表 3 に示す。

Factoid 型と How 型の質問において、質問文をキーボードで直接入力した (音声認識誤りが起きない) 場合と、音声により入力した場合の検索結果を、上位 3 位までの MRR で比較すると、表 4 のようになった。

取扱説明書を検索する場合、質問に対する正しい回答は 1 件しかないことが多く、検索結果画面では 1 位のみ内容を表示している (図 4, 図 5)。よって評価指標として先頭正解を重視する MRR を用いた。また、検索結果の 3 位までを次候補として表示しているため、Factoid 型の質問に対しては上位 3 位までに回答そのものが含まれている場合に正解とし、How 型の質問に対しては正解となる映像、あるいは取扱説明書

MRR (Mean Reciprocal Rank) 正解が最初に出現した順位の逆数を得点としたもので、全質問にわたって平均したもの⁵⁾

のページが上位 3 位までに含まれている場合を正解とした。ただし、それぞれの質問の型が正しく判定されていることも正解の条件とした。たとえば、Factoid 型の質問に対して、How 型と誤って判定され、映像や取扱説明書のページが表示された場合、たとえその中に Factoid 型質問の正解情報が含まれていたとしても不正解とした。

表 4 の結果では、How 型の質問での音声入力による精度劣化が、音声認識精度と比較して小さく抑えられているのに対し、Factoid 型の質問での精度劣化は音声認識精度より著しく大きくなっている(表 4 の(3))。

How 型の質問で呼び出されるビデオ検索エンジンは、ベクトル空間モデルをベースとした文書検索を行っており、質問文中に検索タームが複数含まれているため、1~2 語で音声認識誤りが起きたとしても、その影響を受けにくい。キーボード入力で正しく How 型と判定された質問文の中で、音声入力により Factoid 型と誤判定された質問文は 1 件もなかった。

一方、Factoid 型の質問を処理する質問応答エンジン ASKMi は、質問タイプ/回答タイプ判定、根拠文書検索、回答候補スコアリングの 3 段階で処理している。質問タイプとは Factoid 型や How 型など質問の種類を指す。回答タイプとは Factoid 型の質問に対する回答の種類(重さ、長さなど)のことである。質問タイプ/回答タイプ判定は質問文に含まれる疑問表現に基づいて正規表現で記述されたルールベースで判定する。根拠文書とは、Factoid 型質問の回答が含まれる文書であり、質問中の検索タームに基づいて取扱説明書のページが検索される。回答候補スコアリングは、根拠文書の上位検索結果の文書について、その中の回答候補を検索タームとの距離などに基づいてスコアリングし、結果を出力する。ここで、もし疑問表現の音声認識で誤りが発生すると、質問タイプ/回答タイプ判定が正しく行われず、正しい検索結果を出力することができなくなる。

質問セットをキーボード入力した 96 件と 10 人の被験者により音声入力した 960 件の質問文のそれぞれについて、ASKMi の各処理フェーズの精度を測定したところ表 5 の結果が得られた。

なお、根拠文書検索の MRR (文書単位) は、検索された根拠文書に Factoid 型質問の正解が含まれている場合に、正解として算出した。

表 5 の結果では、(1) の質問タイプ判定の精度劣化が全体システムの精度劣化につながっていることが示されている。すなわちこれは、質問タイプ判定に用いる質問文の疑問表現が音声認識の際に誤認識されてい

表 5 質問応答検索の各処理フェーズにおける精度
Table 5 Accuracy in each step of question answering.

	(1)	(2)	(3)
キーボード入力	1.000	1.000	95.3%
音声入力	0.700	0.964	67.1%

(1)質問タイプ判定精度:

Factoid 型の質問が正しく Factoid 型と判定された割合

(2)回答タイプ判定精度:

Factoid 型と正しく判定されたうち、回答タイプも正しく判定された割合

(3)質問タイプが正しく判定されたと仮定した場合の

全 Factoid 型質問の回答と根拠文書検索の MRR 比:

最終結果の MRR (回答単位) / 根拠文書検索の MRR (文書単位)

ることが全体の検索精度劣化の主要因となっていることを示している。(2) ではキーボード入力と音声入力の両方ともあまり精度が劣化していない。これは、質問文中の疑問表現がもし正しく認識され、質問タイプ判定が正しく行われると、疑問表現に基づく回答タイプ判定も正しく行われることを示している。

表 5 の(3) は質問タイプ判定が正しいと仮定して、すべての Factoid 型質問(キーボード入力 40 件、音声入力 400 件)で質問応答検索をした場合の、最終結果の MRR (回答単位) と根拠文書検索の MRR (文書単位) 比である。キーボード入力に比べて音声入力の場合の MRR 比が著しく低下している。つまり、質問文が音声認識誤りを含む場合、根拠文書検索の精度劣化よりも、根拠文書から Factoid 型の回答を抽出する段階での失敗が最終結果の MRR の精度劣化につながっていることを示している。疑問表現以外の複数の検索タームをもとにする根拠文書検索は音声認識誤りの影響をあまり受けず、Factoid 型の回答を抽出するときの回答タイプ判定は疑問表現をもとにするため音声認識誤りの影響を強く受けるからである。

さらに、音声認識誤りは検索タームよりも疑問表現で頻繁に発生する。これは音声認識システムでの単語の学習が原因だと考えられる。音声認識システムの単語学習には取扱説明書を用いており、検索タームの認識精度は向上するが、疑問表現の認識精度は向上しない。疑問表現が取扱説明書にほとんど現れないからである。たとえば、取扱説明書内に「100 グラム」という表現は現れるが、「何グラム」という表現は現れない。音声認識システムは検索タームに比べて疑問表現が学習されていないため疑問表現の認識で失敗しやすいと考えられる。

表 6 疑問表現の例
Table 6 Example of key expressions.

何回	何グラム	何枚
何倍	何秒	何分
何時間	何週間	何年
何ミリ	何センチ	何ピクセル

音声認識後の質問文では、Factoid 型質問 400 件中 131 件において疑問表現が誤認識されていた。疑問表現の音声認識精度を向上させるためには、疑問表現を用いて音声認識エンジンの言語モデルを学習させる手法がある。しかし疑問表現を多く含むコーパスを得ることは困難であるため、疑問表現の学習は難しい。秋葉ら²⁷⁾は、音声認識で疑問表現（定型表現）の認識率を向上させるため、新聞記事コーパスから疑問表現が含まれる部分を抽出し、疑問表現だけを部分的に強調するように言語モデル（N-gram）を作成する手法を提案している。しかし、我々が用いる検索対象は取扱説明書であり、疑問表現がほとんど含まれないため、秋葉らの手法を用いることはできない。また、表 2 のルールから後述する方法で抽出された表 6 の疑問表現 113 語を用いて音声認識エンジンを学習させた場合、音声認識精度は 0.806 となり、学習前の 0.791 からあまり向上しなかった。これは、文での学習ではなく単語単位でしか学習できないためだと考えられる。よって本実験では疑問表現による学習は行わなかった。

5. 提案手法

前章での音声認識誤りの影響を検討した結果、質問文を音声入力した場合、疑問表現の音声認識誤りが全体の検索精度の劣化につながっていることが分かった。ただし、疑問表現が誤認識されても、音声認識という処理の性質上、類似の発話表現になっている可能性が高いという仮説のもとに、以下の方針で失われた疑問表現の復元を試みることにした。

ステップ 1 あらかじめ疑問表現の候補をリストアップしておく。

ステップ 2 Factoid 型に判定されなかったすべての質問の認識結果において、ステップ 1 の候補リストを取扱説明書での検索タームとの共起関係をもとに絞り込む。

ステップ 3 絞り込まれた疑問表現の候補と、音声認識された質問文の一部との発話上の編集距離を計算し、ある一定の閾値以内のものがあれば編集距離最短の疑問表現を元の質問文に追加する。

処理の例を図 6 に示す。この例では、質問文中の正

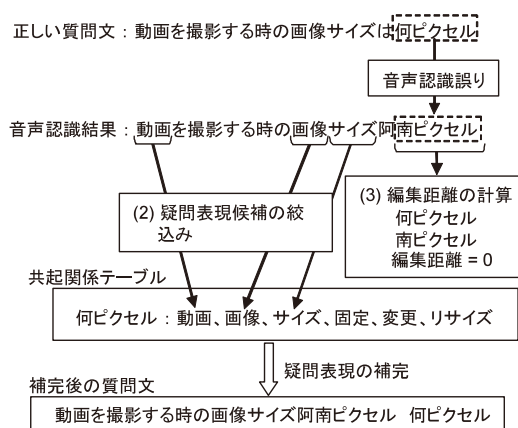


図 6 疑問表現の補充

Fig. 6 Example of a complement of key expression.

しい疑問表現「何ピクセル」が、音声認識誤りによって「南ピクセル」と変化している。そのため、疑問表現をもとに判定している質問タイプ判定が、この質問文を正しく Factoid 型と判定できない。入力された質問文のうち、Factoid 型と判定されなかったすべての質問文で提案手法による疑問表現の補充を試みる。まずステップ 1 で、疑問表現リストをあらかじめリストアップしておく。次にステップ 2 で、疑問表現と検索タームの共起関係テーブルを用いて、質問文中の検索ターム「動画」「画像」「サイズ」のいずれかと共起関係のある疑問表現のみに絞り込み、「何ピクセル」などが疑問表現候補として残る。次にステップ 3 で、絞り込まれた疑問表現候補と、音声認識結果の質問文の一部との発話上の編集距離を計算する。この結果最も編集距離の小さかった「何ピクセル」が選択され、元の質問文に追加され、「動画を撮影するときの画像サイズ阿南ピクセル 何ピクセル」と補充される。

ステップ 1 の疑問表現のリストの一部を表 6 に示す。これは、ASKMi による質問タイプ判定と回答タイプ判定に用いたルール（表 2）から Factoid 型と判定される表現を抽出したものである。今回の実験ではこのような疑問表現を 113 語リストアップした。

ステップ 2 の疑問表現候補の絞り込みで用いる共起関係テーブルは、次のようにして作成する。

ステップ 2-1 検索対象の取扱説明書に対し固有表現抽出を行い、Factoid 型質問の回答候補となる固有表現を同定する。

ステップ 2-2 それぞれの回答候補を質問する疑問表現を推定する。

ステップ 2-3 回答候補の前後一定文字数の範囲に現れるすべての検索ターム（たとえば形態素解析結

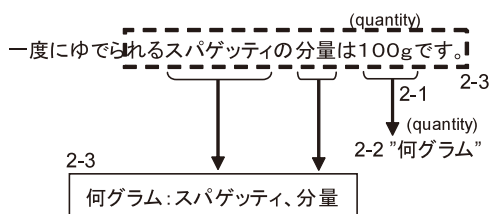


図 7 共起関係テーブル作成処理の例

Fig. 7 Example of collocation table creation procedure.

表 7 共起テーブルの例

Table 7 Example of collocation table.

何回	連続,動き,被写体,赤目,軽減,フラッシュ,発光,...
何倍	ズーム,被写体,光学,デジタル,距離,最大,拡大,画素,...
何枚	ソフトウェア,ケーブル,撮影,間隔,枚数,再生,モード,...
何杯	一度,分量,解凍,キー,中華,バター,ロール,調理,パン,...
何秒	撮影,間隔,枚数,ビュー,画像,表示,プレビュー,発光,...

果の名詞句など)を抽出し,ステップ2-2の疑問表現とともに共起関係テーブルに登録する。

処理の例を図7に示す

図7の例では,ステップ2-1において検索対象の取扱説明書に対し固有表現抽出を行い,回答候補として文字列「100g」とそのラベル「quantity」を抽出する。次にステップ2-2でラベル「quantity」を質問する疑問表現「何グラム」を選択する。次にステップ2-3で回答候補「100g」の前後一定文字以内にある区間「れるスパゲッティの分量は100gです」を切り出し,この部分文字列中に含まれる検索タームとなる語「スパゲッティ」「分量」を抽出し,選択した疑問表現との組で共起関係テーブルに登録する。

ステップ2-1の回答候補は「〇〇cm」のような単位のついた数量表現について行い,またステップ2-2での疑問表現の推定は,その単位の数量を求める疑問表現とするため,同じ固有表現のラベルを持つ疑問表現とした。たとえば,回答候補が「〇〇cm」であれば「何センチ」にした。この処理により「何センチ」などの疑問表現と検索タームとの対からなる共起関係テーブルが作成される。

このようにして作成した共起関係テーブルの例を表7に示す。今回の実験では113語の疑問表現と,それぞれ最大49語の共起する検索タームからなる共起関係テーブルを作成した。96件の質問セットにおい

て,共起関係テーブルの検索タームは,1質問あたり平均2.4語が共起する。

すでに述べたように,ステップ3の発話上の編集距離の計算は,検索タームとの共起関係で絞り込んだ疑問表現の候補リストで計算する。この絞り込みには2つの目的がある。1つは効率化のためである。疑問表現の候補の中には,たとえば「何度」のような短い語もあり,このように短い語でステップ3の編集距離による照合を行うと,多くの語にマッチしてしまう。よって時間短縮のために絞り込みを行う。2つ目の目的は,同音異義語の解消のためである。「何度」「何℃」のように発音が同じで意味の違う疑問表現について,共起関係テーブルを用いた絞り込みを行うことによって,その曖昧性を解消することができる。なお,「長さは?」や「どのくらい」などの特定の単位に依存しない疑問表現は,質問文中の検索タームにかかわらず,すべての質問に対してステップ3の照合処理を行っている。

音声認識によって誤認識された語は,正しい語と表記は違うが発音は類似している。よってステップ3における疑問表現候補と質問文との発話上の編集距離は,音素に近く,疑問表現から容易に作成できるローマ字表記を用いて計算した。また疑問表現候補と照合するのは質問文全体ではなく,質問文中で編集距離が最短の部分文字列を探している。さらに,編集距離の値をそのまま用いると文字数の多い疑問表現が不利になるので,式(2)で正規化して類似度を求めた。

$$1 - D_r / N_r \quad (2)$$

ここで, D_r は疑問表現の候補と質問文中の部分文字列の編集距離(ローマ字表記)であり, N_r は疑問表現の候補のローマ字表記での文字数である。

6. 評価実験

3章での音声認識誤りの影響を検討した実験環境に基づき,5章の提案手法による補完前と補完後のMRR値を比較した。表8は,音声認識誤りのないキーボード入力による精度(1)と,音声入力による補完前の精度(2)と,音声入力による補完後の精度(3)との比較である。表8の結果を図8に示す。この結果,提案手法による質問文の補完により全体の検索精度であるMRR値が0.429から0.597に向上している。

表8の(2)と(3)を比較すると,How型の質問の精度が若干低下している。提案手法による疑問表現の補完はFactoid型と判定されなかった(How型と判定された)すべての質問文に対して試みられるため,(2)で正しくHow型と判定されていても,補完処理により誤った疑問表現が補完されFactoid型と誤判定される

表 8 補完後の検索精度

Table 8 Retrieval performance by complemented query.

	(1)	(2)	(3)
Factoid 型	0.742	0.429	0.597
How 型	0.732	0.625	0.621

- (1) キーボード入力
(2) 音声入力
(3) 音声入力を補完

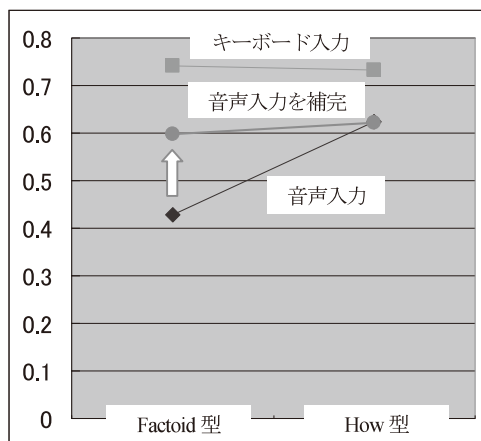


図 8 実験結果

Fig. 8 Result of experiments.

悪影響がある。ただし、このような失敗は (2) で正しく How 型と判定されていた 537 件のうち 5 件にとどまった。さらにこの 5 件のうち 2 件は、(2) でも音声認識誤りの影響で正しく正解が検索されておらず、疑問表現補完による How 型への影響に優位な差はなかった。

質問応答検索の各処理フェーズの精度を調査すると、表 9 の結果が得られた。補完後の (3) の値が音声入力では 0.671 であったのが、補完後は 0.926 まで向上している。これは根拠文書の検索が成功しても、Factoid 型の回答を抽出する段階で音声認識誤りの影響を大きく受けていたのが、補完処理によって質問タイプ判定/回答タイプ判定の精度が著しく向上していることを示している。

次に、提案した補完手法による失敗分析を行う。実験に用いた Factoid 型質問文の実験結果の内訳を表 10 に示す。

Factoid 型質問文 400 件のうち、音声認識誤りによりキーボード入力に比べて精度劣化が起きていた 152 件のうち、35 件は疑問表現が正しく音声認識されており、正しく Factoid 型の質問と判定されているため、本手法による補完処理が行われなかった。つまり本提案手法で対象となるのは疑問表現が音声認識誤りの影

表 9 補完後の質問応答検索の各処理フェーズにおける精度

Table 9 Accuracy in each step of question answering by complemented query.

	(1)	(2)	(3)
キーボード入力	1.000	1.000	95.3%
音声入力	0.700	0.964	67.1%
音声入力を補完	0.923	0.959	92.6%

(1) 質問タイプ判定:

Factoid 型質問が正しく Factoid 型と判定された割合

(2) 回答タイプ判定:

Factoid 型と正しく判定されたうち、回答タイプも正しく判定された割合

(3) 質問タイプが正しく判定された場合と仮定した場合の全 Factoid 型質問の回答と根拠文書検索の MRR 比: 最終結果の MRR (回答単位) / 根拠文書検索の MRR (文書単位)

表 10 実験結果の内訳

Table 10 Analysis of experimental result.

Factoid 型質問 400 件	音声認識誤りの影響でキーボード入力に比べて精度が低下 152 件	疑問表現で音声認識誤りが発生 117 件	提案手法で改善 72 件
		疑問表現以外の原因 35 件	提案手法で改善しなかった 45 件
	キーボード入力の結果と同じ 248 件		

響を受けた 117 件である。

疑問表現における音声認識誤りにより精度劣化が起きていた 117 件の質問のうち、72 件では本提案手法で改善された。残りの 45 件の失敗原因を以下のように 3 種類に分類した (図 9)。

○ 28 件: 何も補完しない

疑問表現だけでなく、検索タームも誤認識されており、疑問表現の補完に失敗し、何も補完しなかった。

例: 入力質問文「デジタルズームは何倍ですか」
音声認識後「デジタル図の穴もうまいですか」

○ 9 件: 補完に失敗

疑問表現の補完は行ったが、間違った疑問表現を補完した。

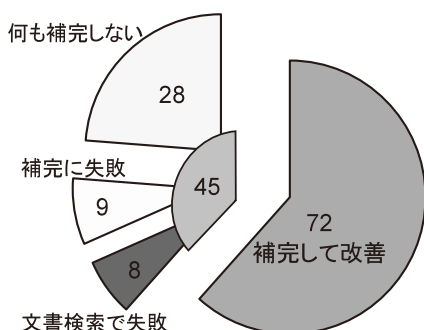


図 9 失敗原因の分類

Fig. 9 Classification of failures.

例：入力質問文「1回でゆでられるパスタは何グラム」

補完後「1回で譲られるパスタは玩具ラグ何分」

○ 8 件：文書検索で失敗

音声認識誤りが起きている疑問表現を正しい疑問表現で補完できているが、根拠文書を検索するフェーズで失敗している。

例：入力質問文「生解凍できる分量は何グラムまで」

補完後「生解凍できる分量判断グラムまでに何グラム」

7. おわりに

本論文では、音声入力を備えた質問応答型マルチモーダルヘルプシステムにおいて、音声認識誤りに対するロバスト性の向上のために、質問を補完する手法を提案し、その効果を実験により確かめた。音声認識誤りによる精度劣化の大きかった Factoid 型質問に対して、検索対象である取扱説明書中の回答候補と検索タームの共起関係に基づく補完処理を実装した。

Factoid 型質問において、疑問表現に基づく質問タイプ判定の精度が、キーボード入力に比べて音声入力の場合は 70.0% まで劣化し、全体の検索精度を劣化させる原因となっていたが、提案手法による補完処理を施すことで、キーボード入力に比べて 92.3% にまで復元することができた。これにより、全体の MRR が 0.429 から 0.597 にまで向上した。

今回、より簡単な操作性を実現するために、質問応答技術に音声入力インタフェースを持たせたが、単に音声認識と情報検索を組み合わせるだけでは、それぞれの精度を単純に結合した精度に到達せず、本論文のような工夫を行うことで初めて 1 つのシステムとして連携が可能となることが分かった。今回は音声認識結

果を情報検索側で補完する手法をとったが、今後、補完結果を音声認識側にフィードバックして認識精度を向上させたり、また対話処理を組み込んでユーザに確認させたりするなど、これらの技術をより密に連携させることで全体の精度を向上させる工夫を検討したい。

参考文献

- 1) Barnett, J., Anderson, S., Broglio, J., Singh, M., Hudson, R. and Kuo, S.W.: Experiments in Spoken Queries for Document Retrieval, *Proc. Eurospeech97*, pp.1323–1326 (1997).
- 2) Crestani, F.: Word recognition errors and relevance feedback in spoken query processing, *Proc. 4th International Conference on Flexible Query Answering Systems*, pp.267–281 (2000).
- 3) Fujii, A., Itou, K. and Ishikawa, T.: Speech-Drive Text Retrieval: Using Target IR Collections for Statistical Language Model Adaptation in Speech Recognition, *Information Retrieval Techniques for Speech Applications*, LNCS 2273, pp.94–104 (2002).
- 4) 伊藤克亘, 藤井 敦, 石川徹也: 未知語検出を用いた語彙統制のない音声検索, 日本音響学会講演論文集, pp.103–104 (2002).
- 5) Fukumoto, J., Kato, T. and Masui, F.: Question Answering Challenge (QAC-1): An Evaluation of QA Tasks at the NTCIR Workshop 3, *Proc. AAAI Spring Symposium: New Directions in Question Answering*, pp.122–133 (2003).
- 6) Hori, C., Hori, T., Isozaki, H., Maeda, E., Katagiri, S. and Furui, S.: Deriving Disambiguous Queries in a Spoken Interactive ODQA System, *Proc. 2003 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.624–627 (2003).
- 7) 堀 智織, 鈴木 潤, 堀 貴明, 磯崎秀樹, 前田英作: 音声インタラクティブ QA のための応答対話文生成法, 日本音響学会秋季講演論文集, Vol.1, pp.81–82 (2002).
- 8) 市村由美, 齋藤佳美, 酒井哲也, 國分智晴, 小山誠: 固有表現抽出と回答タイプ体系が質問応答システムの性能に与える影響, 電子情報学会論文誌 (D-II), Vol.J88-D-II, No.6, pp.1067–1079 (2005).
- 9) Kiyota, Y., Kurohashi, S., Misu, T., Komatani, K. and Kawahara, T.: Dialog Navigator: A Spoken Dialog Q-A System based on Large Text Knowledge Base, *Proc. 41st Annual Meeting of the Association for Computer Linguistics*, pp.149–152 (2003).
- 10) 翠 輝久, 駒谷和範, 清田陽司, 河原達也, 木戸冬子: 音声対話による大規模知識ベース検索システム—音声版ダイアログナビ, *IPSJ SIG Notes*,

- 2004-SLP-53, pp.21–26 (2004).
- 11) Kokubu, T., Sakai, T., Saito, Y., Tsutsui, H., Manabe, T., Koyama, M. and Fujii, H.: The Relationship between Answer Ranking and User Satisfaction in a Question Answering System, *Proc. NTCIR-5 Workshop Meeting*, pp.537–544 (2005).
 - 12) LaLaVoice2001. http://www3.toshiba.co.jp/pc/lalavoce/index_j.htm
 - 13) Magnini, B., Giampiccolo, D., Forner, P., Ayache, C., Jijkoun, V., Osenova, P., Penas, A., Rocha, P., Sacaleanu, B. and Sutcliffe, R.: Overview of the CLEF 2006 Multilingual Question Answering Track, *Working Notes for the CLEF 2006 Workshop* (2006).
 - 14) Masai, Y., Tanaka, S. and Nitta, T.: Speaker-independent keyword recognition based on SMQ/HMM, *Proc. International Conference on Spoken Language Processing*, pp.619–622 (1992).
 - 15) MPEG7. <http://www.itscj.ipsj.or.jp/mpeg7/>
 - 16) Nishizaki, H. and Nakagawa, S.: A System for Retrieving Broadcast News Speech Documents Using Voice Input Keywords and Similarity between Words, *Proc. ICSLP2000*, Vol.3, pp.1073–1076 (2000).
 - 17) 西崎博光, 中川聖一: 音声キーワードによるニュース音声データベース検索手法, *IPSJ Journal*, Vol.42, No.12, pp.3173–3184 (2001).
 - 18) 西崎博光, 中川聖一: 文音声によるクエリーを用いた音声文書検索システムの試作, *IPSJ SIG Notes*, 2002-SLP-41, pp.25–30 (2002).
 - 19) Nitta, T. and Kawamura, A.: Designing a reduced feature-vector set for speech recognition by using KL/GPD competitive training, *Proc. 7th European Conference on Speech Communication and Technology*, pp.2107–2110 (1997).
 - 20) Sakai, T., Saito, Y., Ichimura, Y., Koyama, M., Kokubu, T. and Manabe, T.: ASKMi: A Japanese Question Answering System based on Semantic Role Analysis, *Proc. RIAO 2004*, pp.215–231 (2004).
 - 21) Suzuki, M., Manabe, T., Sumita, K. and Nakayama, Y.: Customer Support Operation with a Knowledge Sharing System KIDS: An Approach based on Information Extraction and Text Structurization, *Proc. SCI 2001*, Vol.7, pp.89–96 (2001).
 - 22) 浦田耕二, 福井美佳, 藤井寛子, 鈴木 優, 酒井哲也, 齋藤佳美, 市村由美, 佐々木寛: 質問応答技術に基づくマルチモーダルヘルプシステム, *IPSJ SIG Notes*, 2004-FI-74, pp.23–29 (2004).
 - 23) Voorhees, E.M.: Overview of the TREC 2004 Question Answering Track, *Proc. 13th Text Retrieval Conference (TREC 2004)* (2005).
 - 24) 筒井秀樹, 真鍋俊彦, 福井美佳, 藤井寛子, 浦田耕二, 酒井哲也: 質問応答型検索の音声認識誤りに対するロバスト性向上, *IPSJ SIG Notes*, 2005-NL-166, pp.31–38 (2005).
 - 25) Tsutsui, H., Manabe, T., Fukui, M., Sakai, T., Fujii, H. and Urata, K.: Improving the Robustness to Recognition Errors in Speech Input Question Answering, *Proc. 3rd Asia Information Retrieval Symposium (AIRS 2006)*, LNCS 4182, pp.297–312 (2006).
 - 26) 河原達也: 話し言葉による音声対話システム, *IPSJ Journal*, Vol.45, No.10, pp.1027–1031 (2004).
 - 27) 秋葉友良, 伊藤克亘, 藤井 敦: N-gram の部分的強調による定型表現への言語モデル適応手法, *電子情報通信学会論文誌 (D-II)*, Vol.J86-D-II, No.12, pp.1727–1736 (2003).

(平成 19 年 4 月 2 日受付)

(平成 19 年 9 月 3 日採録)



筒井 秀樹 (正会員)

1999 年大阪大学大学院工学研究科電子制御機械工学専攻修士課程修了。同年 (株) 東芝入社。現在, 同社研究開発センター知識メディアラボラトリー研究主務。自然言語処理, 音声・マルチメディア検索, ナレッジマネジメントシステムの技術開発に従事。



真鍋 俊彦 (正会員)

1983 年東京農工大学大学院工学研究科数理情報工学専攻修士課程修了。同年 (株) 東芝入社。現在, 同社研究開発センター知識メディアラボラトリー主任研究員。自然言語処理, 音声・マルチメディア検索, ナレッジマネジメントシステムの技術開発に従事。



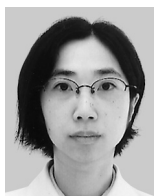
福井 美佳（正会員）

1986 年横浜国立大学工学部情報工学科卒業．同年（株）東芝入社．以後，文書処理，プレゼンテーション，コンピュータ・グラフィックス，ナレッジマネジメント，マルチモーダルナレッジ共有，家電ヘルプ等のシステム開発を通じて，ヒューマンインタフェースの研究に従事．現在，研究開発センター知識メディアラボラトリー主任研究員．1996 年情報処理学会山下記念賞受賞．



酒井 哲也（正会員）

1993 年早稲田大学大学院理工学研究科修士課程修了．同年（株）東芝入社．2000～2001 年英ケンブリッジ大学コンピュータラボラトリ客員研究員．2007 年より（株）ニュースウォッチ自然言語処理研究室室長．FIT2005 論文賞，平成 18 年度本学会山下記念研究賞および論文賞受賞．工学博士．ACM，BCS-IRSG，電子情報通信学会各会員．



藤井 寛子

1994 年慶應義塾大学大学院理工学研究科計算機科学専攻修士課程修了．同年（株）東芝入社．現在，同社研究開発センター知識メディアラボラトリー研究主務．音声・マルチメディア検索，ナレッジマネジメントシステムの研究開発に従事．日本ソフトウェア科学会会員．



浦田 耕二

1992 年（株）東芝入社．現在，同社研究開発センター知識メディアラボラトリー勤務．音声認識，情報検索の開発に従事．