

テレビコンテンツ検索向け音声対話システム

Spoken Dialogue System for TV contents retrieval

伊藤雅弘¹ 溝口祐美子¹ 岩田憲治¹ 若木裕美¹ 永江尚義¹

Masahiro Ito¹, Yumiko Mizoguchi¹, Kenji Iwata¹, Hiromi Wakaki¹, and Hisayoshi Nagae¹

¹ (株) 東芝 研究開発センター 知識メディアラボラトリー

¹Toshiba Corporation, Corporate Research & Development Center, Knowledge Media Laboratory

Abstract: We demonstrate this dialogue system on REGZA Z10X we released in Oct. 2014. This system is allowed to retrieve TV contents such as programs and these scenes through spontaneous speeches and dialogues. One of characteristic points is that it utilizes states of TV such as views and settings to understand a user's intent. Another point is that it has a new words acquisition system to understand frequent-increasing keywords such as TV titles. We demonstrate this dialogue system on REGZA Z10X we released in Oct. 2014.

1 はじめに

近年、テレビ放送の多チャンネル化や、過去数日の全番組を録画できる機器の登場により、視聴できる番組が増大し、ユーザが把握できないほどの番組を視聴できる環境が提供されている。その一方で、見たい番組を探す手段として、リモコンの数字ボタンで検索キーワードを入力する手段しか用意されていない。これでは、たとえ多数の番組が自動録画されていても、ユーザが新しい番組を発見して楽しむことが困難である。そこで我々は、音声対話でテレビコンテンツを検索する対話型インタフェースを開発した。ユーザは自由な文章で番組検索の要望を伝えるだけで、システムはユーザの意図を理解し、テレビと連携しながら適切なキーワードで検索処理を行い、ユーザの要望に合う番組をリストアップする。本システムは、2014年10月に発売されたテレビREGZA Z10Xに搭載されている[1]。本稿では、テレビに搭載されたテレビコンテンツ検索向け音声対話システムの概要について紹介する。

2 機能紹介

本製品では、自由な発話を通して、主に表1のような検索機能を提供することが出来る。また、検索において、表2に示すような検索条件を発話内の表現から識別し受理できる。例えば「TBSで松本人志が出ているシーンが見たい」と発話すると、録画されたTBSの番組中から松本人志が映っているシーンが検索される。

表 1 提供機能

機能	発話例
番組検索	「マッサンが見たい。」「先週の新番組を探して。」
絞り込み検索	「先週ので絞って。」「ドラマに絞って。」
シーン検索	「錦織圭が出てるところを出して。」
番組情報表示	「これって何時に終わるの。」「これのタイトルは。」

表 2 検索条件

検索対象	自動録画番組、手動録画番組、番組表、シーン、YouTube
検索キーワード	番組名、出演者名、コーナー名、ジャンル名、放送局名など

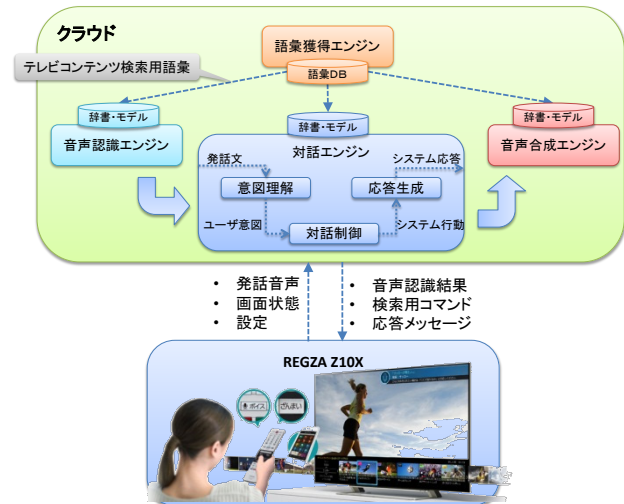


図 1 システム構成

3 音声対話システムの概要

図1に示すように、本システムは、音声認識エンジン、対話エンジン、音声合成エンジン、語彙獲得エンジン、そしてユーザインタフェースを内蔵した

図1のRegza Z10Xに関する画像は[1]の東芝のウェブページから引用しており、著作権は株式会社東芝が有しています。

テレビ (REGZA Z10X) で構成されている。対話エンジンは、これまでに著者らが開発したエンジン[2]を用いており、図に示すように意図理解・対話制御・応答生成で構成されている。

3.1 基本動作

まず、ユーザがマイクに対して発話すると、その発話音声は、音声認識エンジンに送信され発話文が生成される。次に、対話エンジンが発話文・テレビの画面状態 (映像表示中、番組表表示中など)・設定情報 (視聴可能チャンネル、HDD 接続など) を受け取り、まず意図理解部で、ユーザ意図を解析する。さらに対話制御部でユーザ意図からシステム (テレビ) が行うべき動作を決定する。応答生成部は、決定した動作に応じたコマンド (検索、推薦など) 及び応答文を生成する。テレビは受信したコマンドを実行すると共に、音声合成エンジンを用いて生成した応答文に対応する応答音声を再生する。

3.2 対話エンジン

統計的機械学習による意図理解

発話文を解釈しユーザ意図を識別する意図理解部では、統計的機械学習を用いて発話文に対する識別のロバスト性を高めている。そのモデルは、様々なシチュエーションにおける発話表現を集めたコーパスと、そのコーパスに対する自然言語解析の結果を学習させることによって、多様な表現に対応できるようになっている。例えば、シーン検索において発話文に「シーン」という単語を明示的に入れなくても「松本人志が出てるところを出して。」という発話で、シーン検索というユーザ意図として解釈できる。

コンテキスト情報を考慮した対話処理

本システムでは、コンテキスト情報を考慮した対話処理を行っている。1つ目には、絞り込み検索など過去の対話履歴を考慮した対話処理、2つ目には、画面状態などユーザの置かれている状態を考慮した対話処理を行っている。以下で後者に関して具体例とともに説明する。

対話やリモコン操作によって画面状態が変わった後にユーザが発話する時には、画面から分かる情報や操作履歴などを暗黙的なコンテキストとして省略する事が多い。このようなときに、発話文だけを使って対話処理をすると、ユーザの意図しない動作になる可能性がある。そこで、本対話システムではコンテキスト情報を使った対話処理を行っている。ここでは、表3に示した対話例とともにその処理について簡単に説明する。まず、U1でユーザがシーン検索をすると、<シーン検索画面>が表示される。この画面を見ながらさらにシーン検索をしたいと思っ

表 3 対話例

S0	<映像表示>
U1	「松本人志が出てるところを出して。」
S1	「松本人志のシーンを検索します。」<シーン検索画面>
U2	「錦織圭を見せて。」
S2	「錦織圭のシーンを検索します。」<シーン検索画面>

た場合、すで<シーン検索画面>が表示されているため、U2のようにあえてシーン検索を意識した表現をしなくなることがある。もし U2 の発話だけで対話処理をすると、録画番組を見たいのか、シーンを見たいのかなど特定できない。そこで本システムでは、<シーン検索画面>状態で発話を行った場合にはその状態をコンテキスト情報として考慮した対話処理を行い、S2のようにシーン検索として扱えるようにしている。なお本処理は、テレビから受け取った画面状態のほか設定情報も用いることで視聴環境もコンテキストとして扱っている。

3.3 語彙獲得エンジン

本システムは、テレビ番組に関連する番組タイトル・人名など、多数かつ更新頻度の高い固有表現を取り扱う必要がある。特に、番組タイトルは番組改編期ごとにユニークな名称が多く出現するため、それら新規語彙への対応は重要である。図1に示した語彙獲得エンジンは、定期的に番組タイトル・人名などの新たな語彙をシステム上から収集し、さらにクラウドソーシングによる略称・読み付与、重要語彙の絞り込みなどを行い、語彙DBに蓄積する。各エンジンはその共通の語彙DBに基づき、辞書やモデルを再構築することによって、日々増加する固有表現への対応を実現している。

4 おわりに

本稿では、10月に発売されたテレビ REGZA Z10X に搭載された音声対話システムの概要について述べた。本音声対話システムによりユーザは自由な文章で番組検索の要望を伝えるだけで、要望に合う番組を検索することができ、膨大な録画番組を効率よく検索することができる。今後は、ユーザの利用動向を踏まえた性能改善や検索機能の向上を進めていく。

参考文献

- [1] http://www.toshiba.co.jp/regza/lineup/z10x/recording_02.html
- [2] 岩田憲治, 若木裕美, 伊藤雅弘, 山崎智弘, 市村由美, 永江尚義: 課題解決知識を用いた音声アシスタント, 人工知能学会言語・音声理解と対話処理研究会資料, Vol. 67, pp. 13-14, (2013)