テレビの利用スタイルを変える対話処理技術

Spoken Dialogue Technology Allowing Viewers to Efficiently Search for Desired Contents in Recorded TV Programs

向出　隆信　■MUKAIDE Takanobu
河村　聡典　■KAWAMURA Akinori
井澤　秀人　■IZAWA Hidehito

レグザシリーズに搭載された“タイムシフトマシン”機能でテレビ番組を録画することにより，番組を見逃すことなく視聴できるようになった。しかし限られた時間の中で，録画された全ての番組を見ることは困難である。そこで，ユーザーが見たい番組へすばやくたどりつける機能の実現が求められている。

東芝は，より的確で効率的な番組視聴ができるように，ユーザーが見たい番組をすばやく探し出せる“ざんまいスマートアクセス”機能に加え，対話処理技術をテレビに適用してざんまいスマートアクセス機能の利便性を更に高めた“ざんまいスマートアクセスボイス機能”を実現し，4K（3,840×2,160画素）テレビ レグザZ10Xシリーズに搭載した。

The REGZA series liquid crystal display (LCD) TVs equipped with the "time-shift machine" function, which can simultaneously record and temporarily store programs shown on multiple channels without the need for scheduling in advance, offer a new TV viewing style that frees viewers from concern about missing desired programs. In order to allow viewers to easily access contents that they wish to see, functions enabling them to precisely and efficiently search for contents of interest in recorded programs are required.

Toshiba has developed two new functions for this purpose: the "Zanmai Smart Access" function, which allows viewers to rapidly find desired contents; and the "Zanmai Smart Access Voice" function, which provides enhanced usability by applying a spoken dialogue technology. We have released the REGZA Z10X series 4K (3,840 x 2,160 pixels) LCD TVs incorporating these functions.

1　まえがき

東芝は，2014年10月に4Kテレビ レグザZ10Xシリーズを発売した。Z10Xシリーズでは，リアルな精細感や豊かな色再現性を追求した“全面直下LED（発光ダイオード）パネル”の採用や，超解像技術などの高画質化処理で地上・BS（放送衛星）デジタル放送及びブルーレイディスク㈜映像を4K画像に変換する“4Kマスターリファイン”の搭載，4K放送や4K配信への対応など，従来モデルからの更なる高画質化を実現している。また，見たい番組を見逃さない視聴スタイルを提供する“タイムシフトマシン(注1)”にも従来モデルから引き続き対応している。

ここでは，Z10Xシリーズから導入した新機能として，タイムシフトマシンで録画したたくさんの番組の中から見たい番組へすばやくたどりつける“ざんまいスマートアクセス”と，ざんまいスマートアクセスによる番組検索の利便性を飛躍的に高める“ざんまいスマートアクセスボイス機能”について述べる。

2　ざんまいスマートアクセス

ざんまいスマートアクセスは，タイムシフトマシンで録画したたくさんの番組の中から見たい番組をすばやく見つけられるように，旧モデルに導入した機能“ざんまいプレイ”をよりスマートに進化させた機能である。ざんまいプレイの基本思想を損なうことなく，ユーザーの多彩な視聴スタイルを満足できるように，簡単かつわかりやすいGUI（グラフィカルユーザーインタフェース）で好みの番組へすばやくアクセスできる。例えば，番組視聴中にリモコンの“ざんまい”ボタンを押すことで，現在視聴中の番組に関連する番組や，いつも見ている番組，まだ見たことのない番組からのおすすめ番組，新番組，好きなジャンルの番組などを表示させ，選ぶことができる（**図1**）。

また，Z10Xシリーズでは，マイクを内蔵した音声リモコンに向かって話しかけるざんまいスマートアクセスボイス機能に対応することで，検索キーワード入力の煩わしさを解消した（**図2**）。

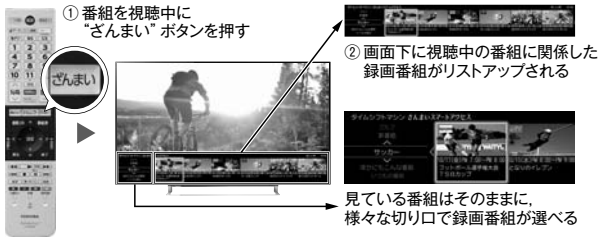


図1．ざんまいスマートアクセス機能の操作フロー ― リモコンのざんまいボタンを押すことで，ざんまいスマートアクセス機能を利用できる。
Operation flow of Zanmai Smart Access function

(注1)　チャンネルと時間を指定して，それに該当する番組を録画して一時保管し，視聴できる機能。

図2. ざんまいスマートアクセスボイス機能の操作フロー ― リモコンの“ボイス”ボタンを押すことで，ざんまいスマートアクセスボイス機能を利用できる。番組選びの指示は声での入力が可能で，ボタンによる選択や決定操作に比べて番組がスピーディに選べる。

Operation flow of Zanmai Smart Access Voice function

3 ざんまいスマートアクセスボイス機能

3.1 ざんまいスマートアクセスボイス機能の概要

ざんまいスマートアクセス機能の利便性を高めるために，テレビ番組検索向け音声対話システムを新たに導入した。このシステムにより，ユーザーは自由な発話を通して，**表1**のような検索を指示できる。また，このシステムは**表2**に示す検索条件を発話内の表現から自動的に理解し検索指示を受理する。例えば「東芝テレビで東芝太郎が出ているシーンが見たい」と発話すると，録画された東芝テレビの番組から東芝太郎が映っているシーンが検索される。

音声対話システムは，クラウドサーバ上に配置された音声認識エンジン，対話エンジン，音声合成エンジン，語彙獲得エンジン，及び音声合成器とユーザーインタフェースを内蔵したテレビから構成されている（**図3**）。

ユーザーがリモコンのマイクに向かって発話すると，その発話音声はテレビを介してクラウドサーバ上に配置された音声認識エンジンに送信され，認識結果としての発話文が生成される。次に，対話エンジンが発話文とテレビの画面状態（映像表示中や番組表表示中など）と設定情報（視聴可能チャンネルやHDD（ハードディスクドライブ）接続など）を受け取り，ユーザーの意図を推定してテレビが行うべき動作を決定し，動作に応じたコマンド，応答文，及び応答文の発音情報を生成しテレビに返送する。テレビは受信したコマンドを実行するとともに，発音情報から音声合成器で生成された応答音声を再生する。語彙獲得エンジンは，テレビ番組の検索に必要な語彙（番組名や出演者名など）を収集し，音声認識や，対話，音声合成用の語彙辞書の作成に使用する。

このシステムにおいて実用的な機能や性能を実現するために，システム構成要素であるエンジンの全てを自社開発している。以下に，音声認識・対話・語彙獲得エンジンについて詳細に説明する。

表1. ざんまいスマートアクセスボイスによる提供機能

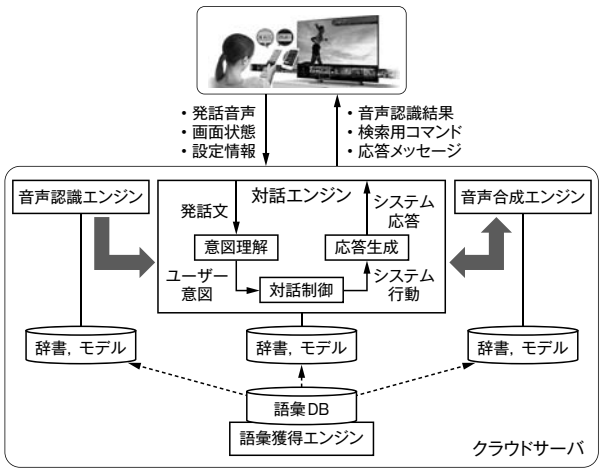
List of functions of spoken dialogue system

機　能	発話例
番組検索	「東芝ドラマが見たい」，「先週の新番組を探して」
絞り込み検索	「ドラマに絞って」
シーン検索	「東芝太郎が出てるところを出して」
番組情報表示	「これのタイトルは」

表2. 検索条件

Search criteria

条　件	内　容
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，コーナ名，ジャンル名，放送局名など



DB：データベース

図3. 音声対話システムの構成 ― 音声認識エンジン，対話エンジン，音声合成エンジン，及び語彙獲得エンジンから構成されるクラウドサービスである。

Configuration of spoken dialogue system

3.2 音声認識エンジン

ユーザーに負担をかけない自由な発話によるテレビ番組の検索を実現するために，あらかじめ定められた表現しか認識できないグラマ（文法）型音声認識ではなく，任意の表現を認識可能な大語彙連続音声認識を用いた。

大語彙音声認識の基本構成を**図4**に示す。認識すべき単語とその発音を記した単語辞書，発音の基本単位である音素の音響的な情報を記した音響モデル，及びある単語に続いて発声されるであろう単語を予測する言語モデルと，音声信号から抽出された特徴を照合することで音声認識結果を出力する。

テレビ番組の検索では，番組名や，出演者名，トレンドワードなど認識すべき語彙の数が非常に多く，発音のよく似た語彙が増加する。また，認識対象が自由な発話なので，不明瞭な発音も多くなるため，音声認識精度の低下が懸念される。

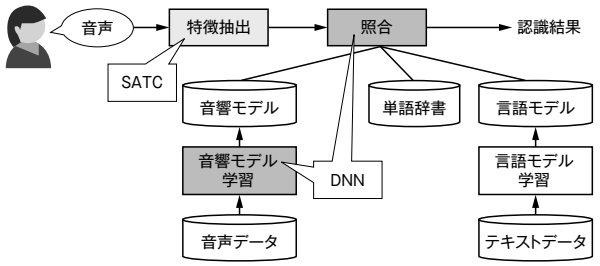


図4．大語彙連続音声認識の構成 — DNNによる音響モデル学習・照合技術と音声特徴量SATCの二つの技術を適用することで，高精度な音声認識を実現した。

Large-vocabulary continuous speech recognition applying deep neural network (DNN) and subband average time cepstrum (SATC)

そこで，図4のように，大語彙連続音声認識の構成要素に対して，DNN（Deep Neural Network）による音響モデル学習・照合技術及び音声特徴量SATC（Sub-Band Average Time Cepstrum）[1, 2]の二つの技術を適用することで高精度な音声認識を実現した。

DNNは層数の深い多層ニューラルネットワークを使用したパターン識別技術である（**図5**）。Deep Learning と呼ばれる機械学習を適用することで，様々な分野において従来の識別器を大きく上回る識別性能が報告されている。今回は，識別しづらい音韻の識別精度向上のために使用した。

SATCは，従来より長い時間の分析窓で切り出した信号における周波数帯域ごとの時間軸上での重心位置情報に基づく特徴量であり，従来の短時間分析窓による振幅スペクトル情報に基づく特徴量と相補的な性質を持つように開発された。従来の特徴量とSATCの併用で，特に会話音声のような発音が不明瞭になりやすい場合の認識性能が向上することがこれまでにも確認されていたが，DNNへの入力として用いることで更なる認識精度の向上を達成している。

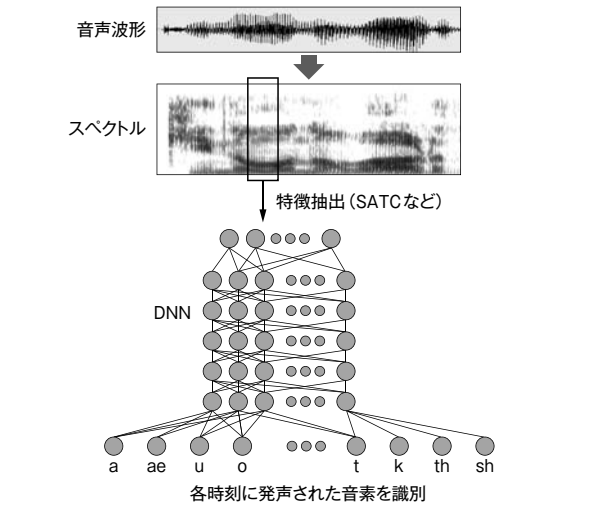


図5．DNNによる音声認識 — Deep Learningと呼ばれる機械学習で，識別しづらい音韻の識別精度を向上させる。

Voice recognition using DNN

3.3　対話エンジン

対話エンジンは，意図理解部，対話制御部，及び応答生成部から構成されている。テレビから発話音声，テレビの画面状態，及び設定情報を受け取り，まず意図理解部で，ユーザーの意図を解析する。次に対話制御部でユーザーの意図からテレビが行うべき動作を決定する。最後に，応答生成部で，システム動作に応じたコマンド（検索や推薦など）及び応答文を生成しテレビに返送する（図3）。

意図理解部では，発話文を解釈してユーザーの意図を識別する。例えば表1に挙げた四つの機能“番組検索”，“絞り込み検索”，“シーン検索”，及び“番組情報表示”のうち，どの機能が実行されることをユーザーが期待しているのかを発話文から識別する。しかし，同じ意図でも様々な発話表現があるため，正しく意図を識別することは難しい。そこで，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（大量の言語データ）とその正解意図ラベル情報に基づいた統計的機械学習により獲得した意図識別器を用いることで，多様な表現の発話文に対する意図理解の精度向上を図っている。精度向上のためには，可能な限り多様な発話表現を収集することが肝要であるが，当社で独自に開発したクラウドソーシングシステム[2-4]を利用することで，現実的なコストで多様な発話表現の収集ができるようになり，高精度な意図理解を実現している。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「東芝花子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ーの意図を解釈できるようになる。

対話制御部では，コンテキスト（前後関係）情報を考慮した対話処理を行っている。(1)過去の対話履歴を考慮した対話処理と，(2)画面状態などユーザーの置かれている状態を考慮した対話処理である。(1)に関しては，通常であれば「ドラマ」と発話すれば「ドラマを検索します」と応答が返り，録画番組からドラマ一覧が検索候補として表示されるが，「東芝テレビの番組が見たい」と発話した後で「ドラマ」と言った場合には，「東芝テレビのドラマを検索します」とする（**表3**）。このように，同じ「ドラマ」という発話であっても，発話履歴を考慮することにより，ユーザーの期待に添ったシステム動作ができる。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ーが発話するときには，画面からわかる情報や操作履歴などの情報は暗黙的なコンテキストとして発話から省略される場合が多い。(2)はこれに対応するための対話処理である。ここでは，**表4**に示した対話例に基づいて，その処理について簡単に説明する。まず，ユーザー1でユーザー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

表3．対話例(1)
Example of spoken dialogue (case 1)

話　者	対話内容
ユーザー１	「東芝テレビの番組が見たいんだけど」
システム１	「キーワードやジャンルをお話し下さい」
ユーザー２	「ドラマ」
システム２	「東芝テレビのドラマを検索します」

表4．対話例(2)
Example of spoken dialogue (case 2)

話　者	対話内容
システム０	＜映像表示＞
ユーザー１	「東芝花子が出てるところ出して」
システム１	「東芝花子のシーンを検索します」＜シーン検索画面＞
ユーザー２	「東芝太郎を見せて」
システム２	「東芝太郎のシーンを検索します」＜シーン検索画面＞

ら更にシーン検索をしたいと思った場合，既にシーン検索画面が表示されているため，ユーザー２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。もしユーザー２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が特定できない。そこでこのシステムでは，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，システム２のようにシーン検索として扱えるようにしている。またこの処理では，テレビから受け取った画面状態のほかに設定情報から得られる視聴環境の情報もコンテキストとしている。

3.4　語彙獲得エンジン

テレビ番組検索向け音声対話システムでは，番組タイトルや出演者名など，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。語彙獲得エンジンは，定期的にこれらの情報をインターネットから取得し，その略称と読みの付与作業をクラウドソーシングに出題し，その結果を語彙DB（データベース）に蓄積する。これらのプロセスは全て自動的に行われ，日々増加する固有表現の音声認識や，対話，音声合成辞書への反映を低コストで実現している。

4　テレビの利用スタイルが変わる

近年，スマートフォンなどのモバイル端末において，音声入力による情報検索が一般的になりつつあるが，テレビに話しかけることに抵抗があるユーザーも少なくない。しかし，これまでの常識であったリモコンボタンによる文字入力と比較しても，音声リモコンによる音声入力は，ユーザーに与えるリモコン操作の煩わしさを大幅に削減した。更に，ふだんの会話のようなあいまいな発話からユーザーの意図を的確にくみ取り，理解する技術をテレビにも適用したことで，スマートフォンにはない価値を与えることができる。Z10Xシリーズでは，これまで述べた個々の技術を複数融合して一つの大きな価値を創造しており，テレビの利用スタイルの変革を体感できる。

5　あとがき

エンターテインメントの多様化により，テレビの視聴時間も大きな影響を受けている。スマートフォンの普及で，多様な情報をインターネットから取得できるようになったことも要因の一つであると考えられる。番組視聴を効率的にする技術をテレビに取り入れて，簡単な操作でユーザーがほんとうに見たい番組にすばやくたどりつける機能を実現できた。

当社が保有する様々な技術を融合して，今後もテレビのユーザーにとって利便性の高い機能を実現していく。

文　献

(1) 中村匡伸 他．“群遅延に基づく音声特徴量の雑音環境下での評価”．日本音響学会2012年春季研究発表会講演論文集．横浜，2012-03，日本音響学会．2012，p.135－136.
(2) 益子貴史 他．同時通訳や音声対話の実用化に向けた大語彙音声認識技術．東芝レビュー．68，9，2013，p.6－9.
(3) 芦川将之 他．“PrivateCrowdSourcingを用いた言語，音声資源の収集～システムの構築と言語収集～”．人工知能学会全国大会（第27回）．富山，2013-06，人工知能学会．2013，3M3-OS-07d2.
(4) 芦川将之 他．Crowd Sourcingを用いた単語への読み付け，アクセント付け手法の提案．電子情報通信学会技術研究報告．AI，人工知能と知識処理．111，447，2012，p.11－16.

向出　隆信　MUKAIDE Takanobu
研究開発センター　ライフスタイルソリューション開発センター　エンベデッドソフトウェア技術開発部グループ長。デジタルプロダクツのソフトウェア開発に従事。
Lifestyle Solutions Development Center

河村　聡典　KAWAMURA Akinori
研究開発統括部　研究開発センター　知識メディアラボラトリー　研究主幹。音声認識，音声処理，及び文字認識の研究・開発に従事。電子情報通信学会，情報処理学会，日本音響学会会員。
Knowledge Media Lab.

井澤　秀人　IZAWA Hidehito
東芝ライフスタイル（株）VS設計統括部 VS設計第五部参事。デジタルプロダクツのソフトウェア開発に従事。
Toshiba Lifestyle Products & Services Corp.