短期間で構築でき業務の自動化・効率化を支える音声対話システム

Rapidly Constructible Spoken Dialog System to Support Automation and Enhance Efficiency of Business Operations

杉浦　千加志 SUGIURA Chikashi

スマートスピーカーなどの，音声対話システムの利用が広がる中，業務の自動化や効率化を目的とした，業務用途での音声対話システムの利用ニーズが高まっている。一方，業務用途で必要になるタスク指向型対話システムでは，発話理解処理の開発にコストが掛かるという問題がある。

東芝デジタルソリューションズ（株）は，音声対話システムを目的ノードと項目リーフで構成されるモデルに構造化し，ノードとリーフに相当する処理モジュールで発話理解を行う技術を開発している。汎用的な設計にすることで再利用もでき，開発コストを削減できる。

As the use of smart speakers and other spoken dialog systems spreads, demand is increasing for spoken dialog systems for business use to automate business operations and enhance their efficiency. However, task-oriented dialog systems required by business applications require costly development associated with the processing of language understanding.

Toshiba Digital Solutions Corporation has structurized a dialog system into a model using purpose-node and item-leaf structures so as to develop a technology that allows processing modules equivalent to a node with leaves to perform language understanding. A generalized design approach has been adopted, making it possible to reuse these modules and thus reduce development costs.

1. まえがき

我が国では，1990年代前半から音声対話システムの研究開発が活発に行われ，東芝グループでもハンバーガーショップでの注文受付を対象に，自由発話を認識し，キーワードスポッティングの手法を用いて実時間で動作する音声対話システムを試作した[1]。2010年代に入ると，スマートフォンで動作する音声対話アプリケーションが，徐々に広がりを見せるようになった。情報検索やネットショッピングなどで音声対話システムの利用が進み，スマートスピーカーの登場をきっかけに，音声対話システムは，より身近な機器操作の手段になってきている。同時に，社内外から多く寄せられる“よくある質問”に対応するスタッフに掛かる，負荷の大きさなどの問題から，業務用途での音声対話システムに対する利用ニーズが高まっている。

業務用途で，音声対話システムを利用する最も分かりやすいシーンは，人が人に音声で目的を伝達して実行する場合に，それを受ける側を機械で自動化するケースが挙げられる。例えば，銀行の窓口業務に適用した例として，当社は，相続相談に関する対話シナリオを備えた“ネット相続相談サービス”を構築し[2]，銀行員の相続相談窓口業務の自動化を実現している。各銀行のサービスに沿って，スムーズな手続きを支援する対話システムは，実店舗での対応業務，及び顧客の来店回数や負担を減らす効果があり，正に働き方改革の一翼を担っている。音声対話システムの利用シーンは，このほかにも，外回りの営業担当者が商品の在庫確認や日報報告をするためにオペレーターに電話をするケースや，機械などの保守点検業務中に異常データを確認した作業員がオペレーターに電話をして対応方法を聞き出すケースなどが挙げられる。

音声対話システムは，ユーザーの課題を解決する“タスク指向型対話システム”と，いわゆる雑談などをする“非タスク指向型対話システム”に大別できる。東芝デジタルソリューションズ（株）は，業務の負荷を軽減するとともに，内閣府が推進している働き方改革に貢献するために，タスク指向型の音声対話システムの研究開発を行っている。ここでは，このタスク指向型の音声対話システムについて述べる。

2. 音声対話システムの課題

2.1 音声対話システムの概要

音声対話システムは，音声認識部，テキスト対話部，音声合成部から構成される。

テキスト対話部の具体例として，当社の音声対話システム[(3)]を**図1**に示す。テキスト対話部は，音声認識部でテキスト化された入力を解析し，発話意図を推定する“発話理解”，ユーザーの発話意図と対話履歴から，次の発話内容を決定する“対話制御”，応答を生成する“応答生成”の三つの基本モジュールから構成される。発話理解には，対話遂行のための知識として，ユーザーの様々な言い回しを理解するための発話理解モデルと，専門的な知識からユーザーのタスクを遂行する対話シナリオがある。

2.2 システム構築における課題

2.1節で述べたとおり，ユーザーの様々な言い回しを理解し，専門的な知識を踏まえて応対する音声対話システムは，業務の自動化に非常に有効に機能する一方で，対話シナリオの構築や変更に要するコストが大きいという問題がある。ユーザーの発話に対して，より細かく意図を理解して適切な対話状態遷移を定義する方が，ユーザーの発話の取りこぼしが減り，自然な対話が可能になる分，シナリオ構築に手間と時間が掛かる。また，これは，対話シナリオの変更についても同様である。3章では，これらの課題の解決に向けた取り組みについて述べる。

3. RECAIUSエージェントプラットフォーム

当社は，業務の自動化や効率化を支援する音声対話システムを“エージェント”と称し，このエージェントを業務用途に応じて容易に構築できるプラットフォームとして，“RECAIUSエージェントプラットフォーム”を開発している。ここでは，コストを抑えるという課題を解決するために開発している，音声対話システムのテキスト対話部について述べる。

3.1 基本対話モデル

対話を小さい単位に分解したものをパーツとしてあらかじめ用意し，このパーツを組み合わせてテキスト対話を行うことで，音声対話システムの構築や仕様変更を容易にする。この場合，汎用的に利用可能なパーツを作るには，工数が掛かる。また，パーツの組み合わせで対話を表現するため，対話シナリオの自由度の面で制約を受ける。しかし，業務用途のタスク指向型対話システムでは会話の内容が限定されるので，これらによる影響は軽微であると考える。

対話の構成要素を汎用パーツ化するために，まず，対話をシンプルなモデルで表現する。人同士の対話を単純化したモデルを“基本対話モデル”として**図2**に示す。依頼者が，対応者に依頼のトリガーとなる発話をし，対応者は，目的達成に必要な各種情報を依頼者から聞き出し，その情報に基づいて目的達成のためのアクションを実行する。依頼者から対応者に対して，目的が明らかに伝えられることもあるが，そうでない場合は目的を聞き出すことも必要である。在庫確認の例では，「在庫確認をしたいです」のように，目的が明確な場合もあるが，「○○商品の件で教えてほしいのですが」などのように，目的が曖昧な場合は，目的自体も含めて情報を聞き出す必要がある。対応者は，情報を聞き出した上で，在庫確認をして依頼者に伝えるなどのアクションを実行する。業務利用の音声対話は，この基本対話モデルで構成されると仮定し，このモデルを並列化・階層化することで多種多様な表現ができる。以下では，簡便のため，並列化構造の基本対話モデルについて述べる。

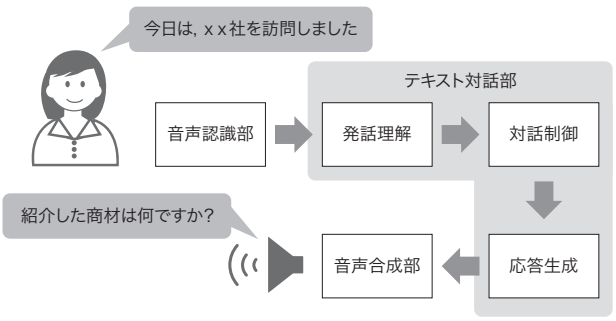

図1. 音声対話システムの概要
一般的な音声対話システムの処理構成を示している。テキスト対話部は，発話理解，対話制御，応答生成の三つの基本モジュールから構成される。
Schematic diagram of spoken dialog system

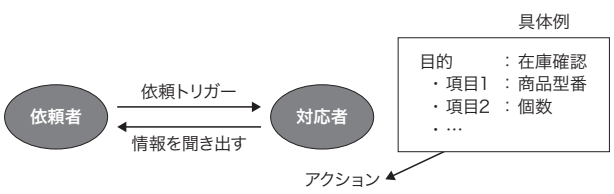

図2. 人同士の対話を単純化した基本対話モデル
依頼者は，対応者に依頼トリガーとなる発話をする。対応者は，情報を依頼者から聞き出し，その情報に基づいて目的達成のためのアクションを実行する。
Simplified basic dialog model of person-to-person dialog

3.2　基本対話モデルの構造

基本対話モデルは，依頼者から目的と目的達成に必要な各種情報（項目）を聞き出してアクションを実行する，というモデルである。これを音声対話システムで実現するために，基本対話モデルと同等の構造を，目的ノードとこれにひも付く項目リーフという形で表現する（**図3**）。ここでは，ユーザーの発話を音声認識した結果のテキストを入力するものとして説明する。

入力テキストは，全ての目的ノードと項目リーフに入力される。各項目リーフ部は，事前に定義された動作定義に応じて，{確定，曖昧，リジェクト}のいずれかを判定する。項目リーフの型が選択肢型の場合，入力テキスト中に選択肢のいずれかがあれば“確定”と判定して項目を特定する。また，類似するものがあれば“曖昧”と判定して問い返しを行い，なければ“リジェクト”と判定して何もしない。日時型の場合も，同様に入力テキスト中に日時表現があれば項目を特定し，曖昧な場合は問い返し，それ以外はリジェクトする。数値型の場合は，入力テキスト中に数値があるだけで項目を特定すると過剰に項目が特定されることが想定されるため，項目名や単位などの情報を加味して項目を特定し，曖昧な場合は問い返し，それ以外はリジェクトする。テキスト型の場合は，項目名が入力テキスト中にあればそれ以降のテキスト内容として項目を特定し，曖昧な場合は問い返し，それ以外はリジェクトする。このように，各項目リーフ部は“型”に応じて適切に振る舞いを変えるように定義する。例えば，入力テキストが「ABC001の在庫は?」の場合，商品型番の選択肢にABC001が存在すれば，項目リーフA1とC2は確定となり，その値はABC001となる。また，そのほかの項目リーフはリジェクトになる。

目的ノードも同様に{確定，曖昧，リジェクト}のいずれかを判定する。入力テキスト中に目的名が完全一致で含まれている場合は，確定と判定して目的を特定する。入力テキスト中に目的名の一部が含まれていたり，下位層の項目リーフに確定若しくは曖昧があったりする場合は，曖昧と判定して問い返しを行い，それ以外はリジェクトする。

例えば，入力テキストが「ABC001の在庫は?」の場合，目的ノードAの目的名「在庫確認」の一部「在庫」が入力テキストに含まれ，また，ABC001によって，ひも付く項目リーフA1が確定なので，目的ノードAは曖昧となり，目的ノードCは項目リーフC2が確定なので曖昧となる。曖昧な場合は問い返すので，目的が在庫確認なのか故障の問い合わせなのかを問い返す。

処理制御部は，問い返しの回数を減らすために，目的ノードの複数が曖昧と判定された場合は，その中でスコア付けして，スコアが一番高いものだけを確認するなどの制御を行う。例えば，在庫確認のスコアが一番高い場合，「在庫確認でよろしいですか?」と応答し，ユーザーが「はい」などの同意を意図する発話をしたら，目的ノードAを確定と判定し，一旦，目的ノードAだけが優位な状態で，そのほかの目的ノードはリジェクトと判定するように制御する。このとき，目的ノードAの項目リーフA2は，値が未定なので，個数を確認する応答をユーザーに問い返し，項目リーフA2は上述の処理仕様に従って値を特定する。

目的ノード下位層の項目リーフが全て確定したら，業務用途ごとに事前に定義された所定のアクション（例えば，在庫確認データベースに問い合わせた結果を出力するなど）を実行する。項目リーフは，選択肢型，数値型，のように型ごとに振る舞いを決めて作ることで汎用性が高まるので，一度作成すれば再利用が可能となる。目的ノードも同様に，項目リーフの値そのものに依存しない振る舞いとすることで汎用

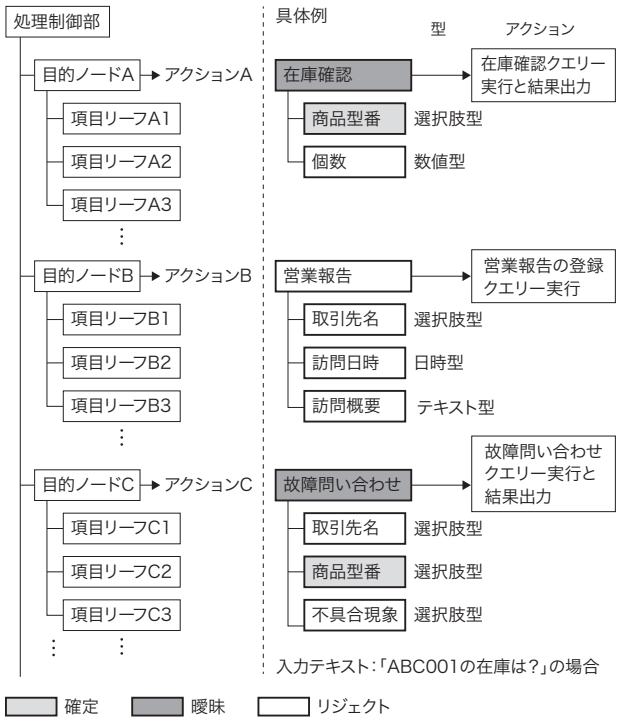

図3．基本対話モデルの構造と具体例

基本対話モデルと同等の構造を，目的ノードと項目リーフで表現することで，音声対話システムの短期間での構築や仕様変更の容易性を実現できる。

Example of configuration of basic dialog model using purpose-node and item-leaf structures

性を担保できるので再利用ができる。

この方法により，型ごとの項目リーフをあらかじめ作成することで，2.2節で示した課題を解決できる。システム要件に応じて目的ノードと項目リーフの構成を定義することが，ユーザー意図の定義と同義であり，目的ノードと項目リーフの判定状態が，対話状態遷移を暗に示すことになる。

この音声対話システムは，システム全体を統括的に管理するものではないため，目的ノードと項目リーフそれぞれの，動作定義並びに処理制御部を，独立性と汎用性を伴うように事前に作り込んでおくことが重要となる。

3.3 開発した技術で期待される効果

⑴ システム構築の容易性向上　音声対話システムで実現したい目的と，その目的ごとにユーザーが入力する必要がある項目を定義するだけで，音声対話システムが，短期間で容易に構築できる。

⑵ 仕様変更への対応容易性向上　目的や項目の追加や削除がある場合，目的ノードと項目リーフの構成を組み替えるだけで対応できるため，音声対話システムの仕様変更への対応が容易になる。

また，用途に応じた特殊ケースへの対応も容易となる。例えば，医療介護の現場におけるバイタル報告の目的で，項目の一つに“体温”があるケースを考える。汎用的に作った数値型の項目リーフでは，音声認識されたテキスト中の数字列をそのまま数値として抽出する場合，「さんじゅうろくてんごど」と発話された場合は，「36.5℃」と認識されるため正しく処理を行えるが，「さんじゅうろくどごぶ」と発話されたものを「36度5分」と認識したり，ユーザーが発話を省略して「ろくどごぶ」と発話したりすると，本来の意図とは異なる値が抽出される。これらは，音声認識の後処理や音声対話システムの後処理などで対応することも可能だが，開発コスト増となる。開発した手法では，このようなケースでも，“体温型”の項目リーフをあらかじめ作成することで，音声対話システムの前後に影響を与えずに，柔軟な対応が可能となる。

⑶ 音声誤認識への対応容易性向上　従来の発話理解処理は，一般には自然言語処理をベースとするため，音声誤認識による影響が大きい。一方，開発した手法では，誤認識による対話精度劣化を低減する手立てを打ちやすい。例えば，選択肢型の項目リーフ“取引先”の選択肢に“東芝(とうしば)”がある場合，ユーザーが「東芝」と発話して仮に“投資が”と誤認識されたときを考える。開発した手法では，取引先の選択肢で“東芝”は選択されないが，音声認識時の読み情報を用いて類似マッチングして選択肢を特定するなどの改善策を，システム全体への影響をほとんど考慮することなく取ることができる。

⑷ 既存の対話シナリオが流用可能　目的ノードと項目リーフは，入力テキストに対して{確定，曖昧，リジェクト}のいずれかを判定し，曖昧な場合は問い返すという共通のインターフェース仕様となっている。したがって，例えば，既存の対話シナリオをこの仕様でラップするように目的ノードの下位に構成することで，過去の対話シナリオなどを流用することも可能になる。

4. あとがき

当社は，シンプルな基本対話モデルの組み合わせで表現される対話モデルと，これを目的ノードと項目リーフで構成し，各目的ノードと項目リーフで発話理解を行う音声対話システムを開発した。基本対話モデルの組み合わせで表現できる範囲の対話システムという制約の下，目的ノードと項目リーフの動作定義を汎用化することで再利用ができるので，音声対話システムの短期間での構築や，構築コストの低減が可能となる。

今後は，目的ノードと項目リーフによる，発話理解の更なる精度向上を行い，報告業務やヘルプデスク業務など業務の自動化や効率化が求められる業務領域での適用を進めることで，音声対話システムの利用ニーズに応えていく。

文　献

⑴ 竹林洋一，ほか．不特定話者音声自由対話システム TOSBURG Ⅱ —マルチモーダル応答と音声応答キャンセルの利用—．電子情報通信学会論文誌A．1994，77，2，p.241–250.
⑵ 東邦銀行．“東邦の相続相談サービス”．東邦銀行．<http://www.tohobank.co.jp/kojin/others/souzokusoudan.html>，(参照 2018-05-21)．
⑶ 岩田憲治，ほか．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．2013，67，p.13–14.

杉浦　千加志　SUGIURA Chikashi
東芝デジタルソリューションズ（株）
RECAIUS 事業推進部 事業開発部
日本音響学会会員
Toshiba Digital Solutions Corp.