

会話からキーワードのみを検出する音声認識技術

事前学習なく高速・高精度な音声検索を実現

カーナビや携帯電話のハンズフリー操作、電話によるサポート窓口の自動音声応答システム等、私たちの生活シーンには音声を用いたサービスが多く提供されていますが、いずれも人が機械にあわせる必要があります。富士通研究所は、人に制約がかからない音声認識技術を目指して、自然な会話からリアルタイムでキーワードを検出する技術や録音データにインデックスを付けて頭出し再生が可能な技術を開発しました。

①ワードスポッティング
ユーザーの発話からキーワードリストに登録した単語のみを抽出し、それ以外の発話は無視する音声認識技術。

自然な会話をサポートする音声認識へ

一般的な音声認識技術は、機械が認識できる言葉をあらかじめ登録して、その言葉だけを人が機械の

タイミングにあわせて発話する必要がありました。富士通研究所は、人にやさしい技術が求められていると考え、こうした制約を排除し、自由な発話を認識できる技術を開発しました。

自然な会話に機械が対応するには、従来、辞書や言語モデル（単語と単語のつながりやすさを確率で表したものを）を構築し、機械に事前学習させた上で、人の会話を全て抽出して文章化する、という手法があります。しかしこの方法は、事前学習にかかる時間と費用が膨大となります。そこで富士通研究

所は、会話の中のキーワードだけを抽出する「ワードスポッティング」^①という音声認識技術に着目しました。この技術であれば、事前学習を必要とせず、ユーザー自身でキーワードリストの作成や変更・削除等のメンテナンスを行うことも可能です。

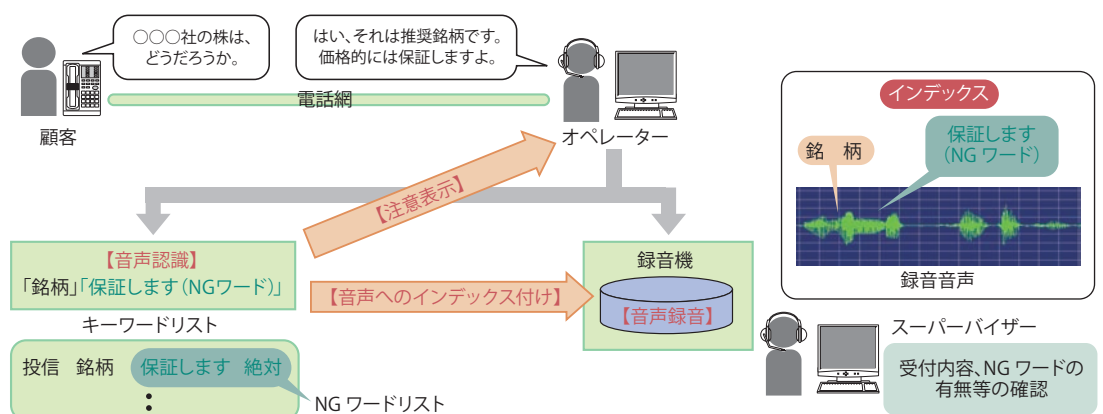
会話や録音データからキーワードを検出

ワードスポッティング技術を応用しコールセンター業務を支援する技術として、次の2つの技術を開発しました。

通話中にキーワードの有無をモニタリングする技術

キーワードリスト内の言葉にマッチングした発話のみを拾い出し、画面にリアルタイム

■ キーワード検出によるコールセンター業務支援



で表示できます。これにより、例えば、証券会社の売買窓口において、オペレーターが顧客に必ず説明しなければならない項目の言葉や、逆に言うてはならないNGワード（例：保証します）の発話状況を、バックヤードのスーパーバイザーがモニタリングするといった対応が可能になります。

録音データからキーワード発話箇所を頭出しする技術

録音データとキーワードリストのマッチングによってキーワードが出現する箇所にインデックスを付け、音声波形のタイムラインに表示できます。これにより、例えば、過去の通話から監視が必要と思われるものをピンポイントで取り出したり、特定の言葉が発話されている部分だけを頭出し再生するといった録音データの操作を、迅速に行うことが可能になります。

類似語除去や不特定話者に対応し高速・高精度を実現

本技術は、次のようなアプリケーションや技術との連携によって高速化・高精度化が図られています。

(1) 類似語除去：システム側でキーワードの類似語を自動生成することにより、例えば「和歌山県」という言葉をキーワードにした場合、音が似ている「岡山県」等を誤検出することを防ぎます。

(2) 不特定話者対応：数千人規模の音声から集めた波形データを基に生成した独自の音響モデル（音の特徴をモデル化したもの）を応用し、不特定話者が発話するキーワードを、電話音質（サンプリング周波数8kHz）

で認識します。

(3) 高速インデックス処理：音声ファイルのインデックス付けを、録音時間の数十分の一の時間で処理します（インテルXeonプロセッサモデルのサーバ利用時）。

(4) 周囲に雑音のある環境下でも音声認識できるよう、雑音波形のモデルを推定して雑音を抑圧したり、雑音は雑音として認識させることで雑音の影響を軽減する等、コールセンターでの利用を想定した雑音対策を行っています。

振り込め詐欺防止に向けた取り組みも開始

本技術は、株式会社アニモ^②の日本初キーワード型音声検索ソフトウェア「VoiceTracking/KeywordFinder（ボイストラッキングキーワードファインダー）」において、NGワードや販促用ワードの検出、FAQ自動表示等のエンジンとして採用され、金融機関・流通・サービスを始めとするコールセンターに導入されています。

また一方では、振り込め詐欺防止に向けた取り組みも始まっています。本技術を、人間の声の調子を検出する技術と組み合わせることで、被害者の異常な心理状態における発話を検出し、注意喚起するための仕組みを国立大学法人名古屋大学と富士通で共同研究^③しています。

今後は、マーケティング活動への適用や、さらにはICレコーダ等の小型端末への搭載も視野に入れ、引き続き速度や精度の向上と機能強化を図っていきます。

②株式会社アニモ
富士通のベンチャー第一号企業として、1994年8月に創業。「音・音声」をキーテクノロジーとしたマルチメディア・ソフト、サービス及びソリューションを開発、販売している。

③共同研究
本研究の一部は、独立行政法人科学技術振興機構（JST）の戦略的創造研究推進事業（CREST）の支援により行っている。