

コンタクトセンター 支援ソフトウェア CSVIEW/VisualVoice V2.3

小此木 崇宏

要旨

コンタクトセンター向け音声認識製品であるCSVIEW/VisualVoiceの概要と、2010年4月に出荷を開始する最新バージョンで追加される機能について紹介します。

キーワード

- コンタクトセンター
- コールセンター
- 音声認識
- 音声入力
- テキスト化
- オペレータ
- 電話音声
- 通話音声

1. はじめに

CSVIEW/VisualVoiceは、コンタクトセンター向けの音声認識ソフトウェアです。オペレータとお客様との電話による会話を音声認識によって自動的にテキスト化することで、業務の効率化やCSの向上などに貢献する製品です。

コンタクトセンターへの音声認識システムの導入に際しては、活用方法や導入効果などのソリューション面での課題のほかに、ソフトウェア製品としての技術的な課題や運用面、導入コスト面での課題があり、その解決のためにさまざまな工夫が必要です。

本稿では、CSVIEW/VisualVoice（以下、VisualVoiceと表記します）の概要と、最新の機能について解説します。

2. 製品概要

音声認識技術には、限られた範囲の単語のみを認識するものや、丁寧な文章を読み上げたような音声を認識するものなどいくつかのレベルがありますが、人と人とが普通に会話をしているときの音声を認識する「会話音声認識」の技術が最も難易度が高く、実用化されている製品も多くありません。

現在の会話音声認識技術は、特定の利用シーンに特化した音声認識用の辞書データを用意することで実用的な音声認識率を実現しています。この辞書データは、単語の情報とその用法（単語どうしの連なりに関する統計情報）とから構成さ

れています。「言語モデル」と呼ばれます。広い範囲の話題に対応できるようにするには、言語モデルに登録する単語やその用法の数が膨大になって認識処理に時間がかかる上に、選択肢が増えることによる認識誤りの可能性が増えてしいます。そのためなるべく話題を限定した方が認識率は高くなります。

その点で、コンタクトセンターの音声は音声認識に適しているといえます。なぜなら、ある特定のコンタクトセンターで交わされる会話の内容は、その企業が取り扱っている製品やサービスに関するものに限定されるからです。

VisualVoiceの最も基本的な機能は、オペレータとお客様の会話音声を認識して自動的にテキストに変換することです。



図1 CSVIEW/VisualVoiceの機能

そしてその機能をベースに、コンタクトセンターの業務に即したいくつかの機能を有しています。

以下にVisualVoiceの主な機能について説明します（図1）。

2.1 オペレータ向け機能

（1）応対メモ（通話メモ）の自動作成

オペレータとお客様の電話による会話音声をオペレータのパソコンに取り込み、その場でテキストデータに変換して画面に表示します。そのデータは保存され、後で参照することも編集することも可能です。テキストだけでなく、音声データも一緒に保存されるので、必要ならば通話の一部を聞き直すこともできます。

（2）検索システムとの連携

通話内容をテキスト化する処理と並行して、あらかじめシステムに登録しておいたキーワードが会話の中に出てきた時点でそのキーワードを抽出することができます。更にそのキーワードを特定の検索システムに受け渡すことができるので、お客様からの質問に対する回答をFAQシステムで探す場合などに、オペレータが自分で検索画面を開いて検索語をキーボードから入力する必要がなくなります。

（3）音声入力機能

お客様との通話ではなく、オペレータが一人で話す内容をテキストに変換することも可能です。具体的には、お客様との通話を終えた後にその応対記録を業務システムに入力する際、キーボードではなく音声で入力を行うなどの用途が考えられます。これにより、キーボードでの入力操作に不慣れなオペレータでも素早くテキスト入力を行い、次の通話を受けるまでの時間を短縮することができます。

2.2 スーパーバイザー向け機能

（1）リアルタイム検知

オペレータとお客様の会話中に特定のキーワードが現れた場合や、ある条件に合致するフレーズが発話された場合に、その場でスーパーバイザーに通知する機能です。緊急対応が必要とされるコールが入ったことを管理者が素早く知ることができたり、オペレータが不適切な発言をしたことを検知して迅速なフォローを行うなどの利用方法が考えられます。

（2）通話データ管理・再生機能

VisualVoiceでは、オペレータとお客様の通話音声とそれをテキスト化したデータとを紐付けてサーバに蓄積することができます。一般的な通話録音装置とは異なり、通話内容をテキスト化したデータが音声と一緒に保存されるので、例えばある特定の製品に関する問合せだけを製品名で検索をかけることにより抽出して内容を確認するといったことを容易に行うことができます。また、再生したい個所をテキスト上で指定することで音声データの該当部分が再生されるため、長い通話であっても必要な個所だけを効率良く再生・確認することができる。

3. 新機能の紹介

VisualVoiceは2009年10月に発表した最新版ver.2.3で「通話録音装置連携」と「言語モデルメンテナンスツール」の2つの新機能を搭載しました。本章では本機能の概要について紹介します。

3.1 通話録音装置連携

既に通話録音装置を導入済のコンタクトセンターの場合、蓄積されている大量の音声データに含まれている価値ある情報を有効に活用したいという強いニーズがあります。しかし、音声データのままでは文字情報としての取り出しは困難であるため、音声認識を使ってテキスト化することが必要になります。上述の通りVisualVoiceは通話音声をリアルタイムにテキスト化する機能は有していますが、更に通話録音装置から録音された音声データを取り込み、テキスト化した内容と音声と連動再生するためのインターフェースを、新規に追加することにしました。

今回の開発では、通話録音装置から音声データを受け取って認識するインターフェース部分の開発と合わせて、圧縮された音声を認識するために必要な「音響モデル」の開発も行いました。「音響モデル」とは、入力された音声データから人の声だけを取り出し、どのような言葉に該当するかを解析していく際に必要となる基礎データ、すなわち「音に関する基本情報」ということになるわけですが、これは元の音声データの品質（雑音の有無やその大きさ、圧縮率やサンプリングレートなど）に合ったものを用意する必要があります。これ

までは、ヘッドセットのマイクから入力されるオペレータの音声と、電話機から取り込むお客様の音声に対応した音響モデルのみを用意していましたが、加えて通話録音装置で保存される一般的な音声データに合わせた音響モデルを用意することにしました。

3.2 言語モデルメンテナントール

VisualVoiceのような音声認識製品にとって、言語モデルを作成・維持するために必要となるコストをいかに削減するかは従来からの大きな課題です。言語モデルは、単語だけを集めた辞書とは異なり、各単語の具体的な用法、すなわち単語どうしの連なりに関する統計データを含んだものです。そのため、ターゲットとする利用シーンで実際に発話される内容を忠実に書き起こしたデータが必要になり、データの準備を含めて多くのコストがかかります。

更に、新しい製品やサービスが出来ば、その名称や関連する表現や言い回しも学習させる必要が出てきます。

言語モデルの作成や更新には非常に専門的な知識とノウハウが必要であり、利用者が日常の中で行えるメンテナントスは新製品の名称など新しい単語を登録することだけでした。それによって新しい単語を認識させることは可能になりますが、その単語の利用法に関する情報までは登録できないため、ほかの単語との組合せによっては認識率に影響が出る場合もあり、これだけでは十分とは言えません。とりわけ、取り扱う業務内容の変化が激しいコンタクトセンターでは言語モデルの定期的な更新が必要になるので、運用コストの増大につながることになります。

そこで、専門的な知識がなくても、利用者自身が言語モデルの更新を簡単に行えるようにする機能が「言語モデルメンテナントール」です（図2）。このツールは、新しい単語を認識させたい場合に、その単語を用いた文章が含まれる文書データを一括登録するだけで、文書内にある単語やその用法、表現などを自動的に学習して言語モデルを更新することができます。したがって、例えば新製品を発表した場合などに用意するマニュアルやFAQ集などのデータを言語モデルに登録することで、その新製品に関する問合せを受けた通話を認識することができるようになります。

単語だけを登録するのではなく、その用法までを文章として登録できることは、単に新しい言葉を認識できるというこ

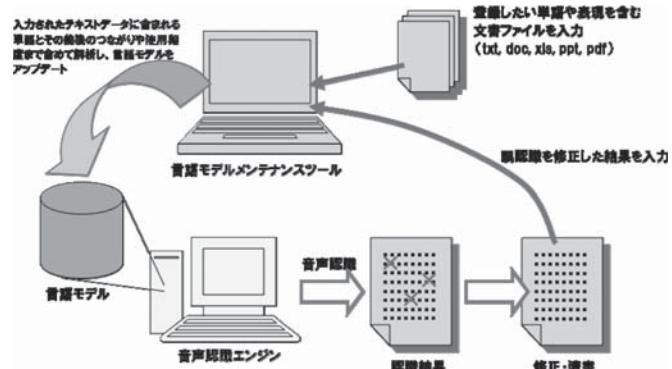


図2 言語モデルメンテナントール

と以上の効果が期待できます。例えば、パソコンに関する問合せを受けるコンタクトセンターで、新たに「リスク」という単語を言語モデルに追加したとします。すると当然、「リスク」という単語が認識できるようになりますが、同時に「ディスク」のような音の似ている別の単語を誤って「リスク」と認識してしまう可能性も出てくることになります。しかし、単語の用法までを情報として学習させることができれば、そのような誤認識を防ぐことができます。人の声には個人差がありますし、ときには正しい発音で話さないこともあります、どんな単語がどのように使われる可能性が高いのかという情報を基に解析すれば、「『ディスク』障害によるデータ損失の『リスク』を回避するために…」のような文章も正しく認識できるのです。

4. おわりに

今回、通話録音装置連携による運用性の向上や言語モデルに関するメンテナントスコスト削減を目的とした機能強化を行いましたが、実際に本製品を利用するオペレータやスーパーバイザーの方々にとって、より使いやすくより役に立つ製品にしていくべきと考えています。今後は、ユーザインターフェースの改良や動作速度の改善など、操作性の向上に関する機能についてもさらなる強化を図っていく所存です。

執筆者プロフィール

小此木 崇宏
第二システムソフトウェア事業部
市場開発推進本部
エンジニアリングマネージャー