

音声認識による検索機能を備えた分散ボイスロギングシステム

藤村卓樹[†] 大島浩太[†] 安藤公彦[†] 寺田松昭[‡]

東京農工大学大学院 工学府[†]

東京農工大学大学院 共生科学技術研究院[‡]

1. はじめに

近年のブロードバンド化や VoIP 関係プロトコルの標準化により、IP 電話の利用者が増加しつつある。コールセンターにおいても通信費の低下に着目し IP 電話システムの導入が進んでいる。また、CRM の一環としてオペレータの教育等の業務改善を目的にコールセンターで音声保存装置を導入し、保存した通話音声の有効活用が行われている[1]。

本研究の目的は、IP 電話環境においてボイスロギングを行い、音声認識を併用したテキストによる音声の検索および参照を可能にするシステムの一方式を提案することである。

2. ボイスロギングシステム

ボイスロギングシステムは、電話等における通話音声を保存するシステムである(図 1)。保存した音声は、音声認識によりテキスト化し、テキストと音声ファイルをマッピングすることにより、テキストを用いたキーワード検索ができる。通話音声の保存は 2 種類の方式が考えられる。(1)保存をサーバで一括して行う方式、(2)通話端末で行うクライアント方式。(1)は、ポートミラーリング機能を備えたスイッチを利用し、音声パケットのコピーをサーバに転送することで行う[2]。(2)は、PC 上で動作するソフトフォンに通話音声の保存モジュールを追加し、保存処理を行う方式である。

本研究では、ロギング処理から保存したデータの検索・参照処理を単一のサーバに依存せず稼働できるシステムにするため、(2)の方式を採用した。しかし、この方式はサーバ方式に比べ、データ保存における堅牢性、データ分散によるアクセスの利便性に欠けるという課題がある。

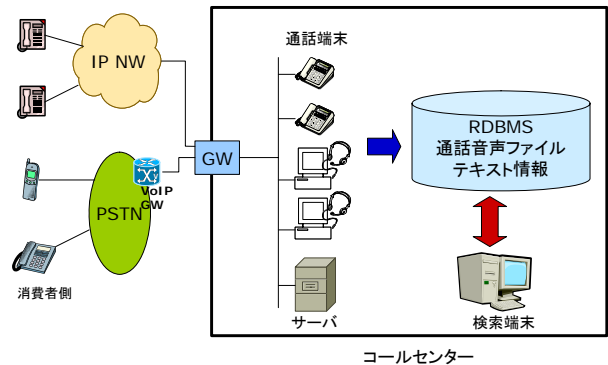


図 1 ボイスロギングシステム概要図

3. 分散型ボイスロギング方式

3.1 ロギング処理

本研究で提案するクライアント型ボイスロギングシステムの構成図を図 2 に示す。同一のスイッチで繋がれた端末同士で一つのセグメントを構成し、このクライアント間で保存音声の分散処理(3.2 節)を行う。

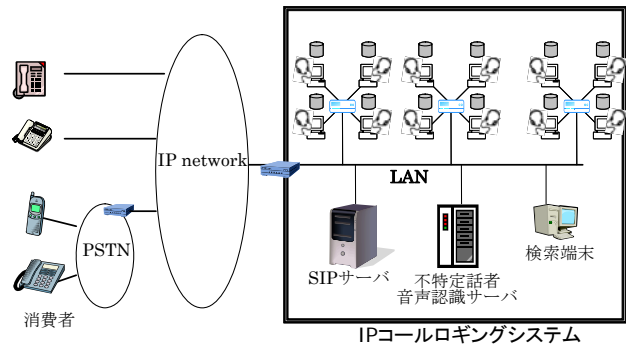


図 2 クライアント型ボイスロギングシステム

通話音声は、通話中に音量をリアルタイムに分析し、無音区間であると判断された部分で区切ることで発声部分のみを抽出し、複数の音声ファイルとして保存する。通話音声は、発呼側と着呼側で別々に保存する。

3.2 分散処理

クライアント方式での音声保存処理は、クライアント機器の故障による保存音声の消失に対応する必要がある。そこで、音声ファイルが無音区

Voice logging system with voice recognition search method
 Takaki Fujimra[†], Kohta Ohshima[†], kimihikoAndo[†], Matsuaki Terada[†]
[†] Graduate school of Technology, Tokyo University of Agriculture & Technology
[‡] Institute of Symbiotic Science and Science and Technology, Tokyo University Agriculture & Technology

間検知により生成されるごとに、他のクライアントに分散保存することにした。時間的に連続する複数の音声ファイルをグループとし、1グループごとに、そのファイル群でビット単位のパリティを計算し、1つの復元補助データを生成する。復元補助データ作成後、音声ファイルとともに他のクライアントに分散する。図3に、ある特定のクライアントが分散保存する例を示す。分散処理を行うことにより、ある1つのクライアントのデータが消失した場合でも、復元補助データと残存ファイルを基に消失したデータの復元が可能である。システム内のすべてのクライアントが同様の分散処理を行うことで、システム全体でデータ消失に対する復元機能を提供できる。

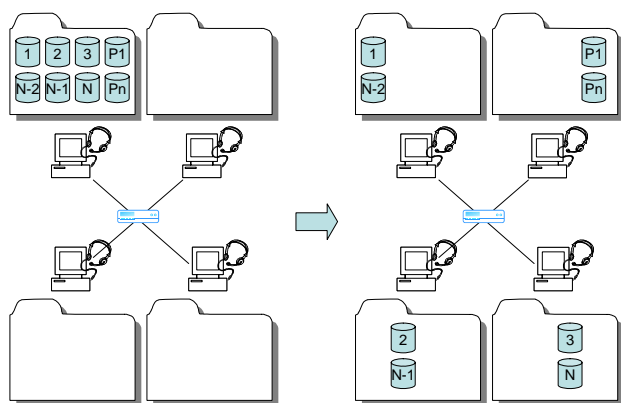


図3 保存音声分散処理の例

3.3 音声認識処理

音声認識は、不特定話者音声認識と特定話者音声認識を併用する。前者には Julius[3]を、後者には Dragon Speech[4]を利用する。使用者が限定されるコールセンター内のクライアント端末に特定話者音声認識を適用することで、通話音声の認識率向上を図った。

3.4 検索・参照

提案した分散処理を行うことで、保存したデータはシステム内のクライアントに分散している。保存音声は検索条件に適合する情報を、分散されたデータの中から参照できなければならない。そのため、各クライアントは、それぞれ DB を備えている。DB に登録する情報にはファイル名、通話を行った時刻、話者、音声認識による通話内容のテキスト、ファイルの分散先がある。検索端末は、各クライアント端末に検索ワードなどの検索キーを送信する。キーを受信したクライアントは自身の保持する DB にアクセスし、キーにマッチするレコードを検索する。該当レコードがあった場合、被検索クライアントは検索端末に結果を返し、

その結果をもとに分散している音声ファイルの保存先が分かり、参照することができる。

4. 実装

プロトタイプシステムの仕様を表1に示す。

表1 実装環境

項目	仕様
OS	Windows XP
シグナリングプロトコル	SIP
トランスポートプロトコル	RTP
RDB	MySQL 4.0.13
Web サーバ	Apache 2.0.46
特定話者音声認識エンジン	Dragon Speech
不特定話者音声認識エンジン	Julius

5. まとめ

クライアントベースのボイスロギングシステムの方式提案を行い、プロトタイプシステムを構築した。ロギング処理において、クライアント間で通話により生成した音声ファイルと、そのファイルから復元補助ファイルを生成して分散することにより、故障等による突然のデータ消失に対する復元性を実現した。また、クライアントごとに DB を保持させ、通話音声のテキスト情報に加えファイル分散先等を併せて登録することにより、分散した通話保存情報を参照することを可能とした。本方式では、通話保存処理から検索まで一貫してサーバレスに行うことができる。

参考文献

- [1] コンピューターテレフォニー編集部・編：コールセンター白書 コールセンター白書 2006、リックテレコム、2006
- [2] 大島浩太、村松英二、大竹八洲孝、大野博樹、糟谷宏、但馬康宏、寺田松昭、“音声情報を含むコミュニケーションの保存・参照技術の開発”、DICO 2004 シンポジウム論文集
- [3] 大語彙連続音声認識システム Julius、京都大学、<http://julius.sourceforge.jp/>
- [4] 音声認識ソフト Dragon Speech、ASCII Solutions、<http://www.dragonspeech.jp/>