

1. はじめに

E-mail はここ数年で広く普及した。また音声を IP で運ぶ VoIP 技術の発展も目覚ましい。本研究では、これらに最近進展の著しい音声認識技術を統合することにより、今までに無い新しい通信サービスを提供することを狙っている。今回は、その第 1 歩として、VoiceMail を音声認識技術でテキストに変換することによって、ユーザーが望む任意のメールアドレスにテキストメールとして配信できるシステムの設計と試作をおこなった。

2. 提案システムの概要

(1) VoIP サービス

本研究室で開発中の SIP (Session Initiation Protocol) サーバーを使いユーザーに VoIP サービスを提供する。

(2) VoiceMail サービス

本システムのユーザーが留守だった場合、本人に代わりメッセージを受け取るサービスを提供する。保存するデータ形式としては MP3 形式を予定している。圧縮率が高く、普及率が高い MP3 形式を用いることによりメールに音声データを添付し配信するサービスも予定している。

(3) Voice To Text Via E-mail サービス

VoiceMail サーバーに蓄えられたメッセージを音声認識技術によりテキストに変換し、テキストメールでユーザーが望む任意のアド

レスに配信することを目的とする。

その際に VoiceMail サービスにより蓄えた音声データを同時添付することにより (4) に挙げるサービスを提供することも考えている。

また本サービスは携帯電話との連携を取ることにより、留守番電話サービスのメッセージ内容をショートメールにより配信できる可能性を持っている。

(4) Voice sync with Text サービス

音声データとテキストデータの同期を取り、音声データにおけるカット&ペーストを提供する。

テキストデータをカット&ペーストすると、音声データ内のテキストに相当する部分もカット&ペーストされる。(図 1)

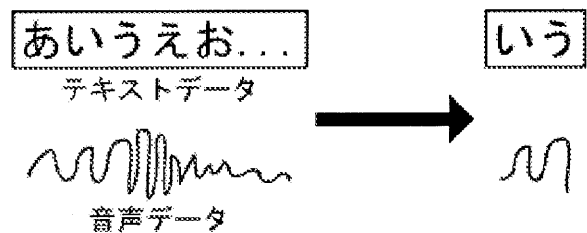


図 1 Voice sync with Text

3. 設計

3.1 方針

VoiceMail をテキストメールとして配信する際の処理の流れを以下に示す。(図 2)

- ① 本システムのユーザーが留守であった場合、電話を SIP サーバーから VoiceMail サーバーに転送する。
- ② VoiceMail サーバーが電話に対応しメッ

セージを保存する。保存する際は MP3 形式で保存する。

③ 保存したメッセージを音声認識技術によりテキストファイルに変換する。

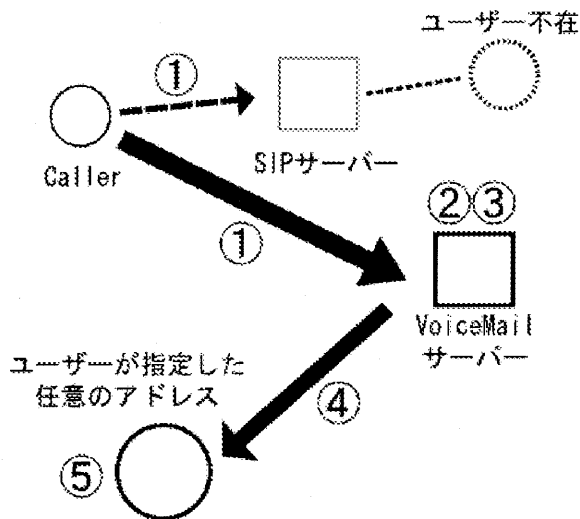


図 2 配信処理の流れ

④ テキストファイルをメール本文として利用し、VoiceMail サーバーによって保存した音声データと一緒にユーザーが望む任意のアドレスに配信する。

⑤ ユーザーは本人が望む任意の場所でメッセージを読むことができ、また音声データによりメッセージを聞くことができる。

3. 2 特徴

① ViaVoice

本システムでは音声認識技術として IBM 社の ViaVoice エンジンを使用する。IBM 社では 1972 年から 30 年余りに渡り ViaVoice エンジン进行研究しており、特定話者の音声認識率は非常に高い。

これを使い VoiceMail サーバーに蓄えた音声データをテキストに変換する。

② Web サーバーとの統合

Web サーバーとのシームレスな統合を図

る。このことにより CGI を使ったサービスを提供し、ユーザーが任意の場所から配信先のアドレスの変更、音声データの添付の有無など様々な設定を行えるようにする。

また Web との統合により任意の場所から VoiceMail サーバー内の音声データ等にアクセスできるようなインターフェースも提供する。

4. 実装

VoIP サーバーとして本研究室で開発している SIP サーバーを配置し、前節に沿った実装を行う。

(1) VoiceMail サーバー

ユーザーが留守であった場合にメッセージを MP3 形式により音声データとし保存すると同時に、音声データをテキストに変換する。

(2) ViaVoice エンジン

VoiceMail サーバー内に ViaVoice エンジンを組み込んだ音声認識プログラムを組み込む。このプログラムを使うことにより、SIP サーバーから転送された電話を受け取った後に自動で音声を変換する。

5. 結言

VoIP サービスと音声認識技術との統合により新しい通信サービスを提案した。このサービスを使うことにより携帯電話を含めた新しいサービスの実現が可能と思われる。

今後の課題として、バックグラウンドノイズの除去、不特定話者への対応が考えられる。