

非同期型音声会議システム VoiceCafe

VoiceCafe: An Asynchronous Voice Meeting System

西本 卓也[†] 北脇 裕康[‡] 高木 治夫^{‡ §}
Takuya Nishimoto Hiroyasu Kitawaki Haruo Takagi

1. はじめに

音声を用いて非同期かつ双方向性のある会話が可能なシステムとして提案された AVM (Asynchronous Voice Meeting) [1] [2] に基づいて、PC、電話、PDA の各クライアントから音声メッセージを共有・操作できるシステム VoiceCafe を開発した [3]。特に設計においては情報アクセシビリティおよびユーザビリティを考慮した。

個人がウェブを通じて手軽に情報交換するためのシステムとして Blog (Weblog) が注目されている。本研究の関連技術として、音声クリップの掲載が可能な Audio Blogging や、携帯電話などのモバイル環境から Blog を更新できる Moblogging などがある [4]。

これらの Blog システムはテキストベースの情報交換を主体にしており、ウェブブラウザで閲覧されることを想定している。これに対して VoiceCafe は、音声による情報交換を主体に設計されており、顧客サポート、障害者支援、高齢者介護、遠隔教育など、幅広い用途や環境でのコミュニケーション支援を目指している。

2. システムの基本設計

VoiceCafe では会議室を「グループ」と呼ぶ。ユーザは各グループ（「公開」または「非公開」）にメンバー登録でき、また自由に新規グループを作成できる。非公開グループへの参加には、グループ管理者の許可が必要である。グループのメンバーが 1 人であればボイスメモとして、2 人であればプライベートな会話の場として、多人数であれば会議の場として利用できる。

各グループにおける発言は「メッセージ」単位で保存される。1 つのメッセージは音声による 1 回の発話であり、肉声そのままデータベースに格納される。

メッセージは「スレッド」単位でツリー状に管理され、親メッセージと子メッセージのリンク情報（相対的なタイミング）が保持される。メッセージの再生中に行なわれた発言操作を、当該メッセージへの返答と見なすことにより、メッセージの引用や参照にあたる操作（親子関係の指定）を効率化する。

各メッセージは「相槌 / 非相槌」属性（発話区間の継続長によって判定）を持つ。スレッド単位でメッセージを再生する場合には、タイミング情報と相槌属性を用いて、なるべく複数のメッセージを重ね合わせて再生する。メッセージを個別に再生する従来のボイスメールと比べて、再生時間が短縮され、仮想会議の臨場感が得られる。

GUI が利用できる場合には、発言された音声メッセージのディクテーション結果をツリー表示してメッセージの閲覧に用いる。

各ユーザがメッセージを再生することにより既読フラグがセットされる。未読メッセージのみを再生することによってメッセージ聴取を効率化できる。

3. 開発における要求

システムの開発において必要と考えた事項を挙げる。

(1) 音声メッセージの継続時間や発話タイミングなどを正確に取得できることを重視し、基本設計の利点を活かした実装を行なう。

(2) 複数の利用環境でデータベースを共有でき、ユーザ数やトラフィックの増加に耐え、クライアント環境を容易に追加できるように、拡張性や柔軟性を持たせる。

(3) 操作が理解されやすく効率的であることを目指す。また、視覚障害者をヘビーユーザと位置付けて、視覚に頼らずに操作できるようにする。ただし PC 環境においては GUI も有効活用する。

4. システム構成

システムを 3 階層（クライアント / ウェブサーバ / バックエンドサーバ）に分割し、通信プロトコルを統一することで、柔軟にクライアントを追加できるように設計した。各階層の役割を以下に述べる。

クライアント：PC 環境、電話環境、PDA 環境の全てにおいて、ユーザの要求は HTTP でウェブサーバに送られ、HTML/XML ファイル、音声ファイルなどがクライアントに返送される。

ウェブサーバ：各クライアントの要求 (HTTP) を受理し、バックエンドサーバに処理を委譲する。

バックエンドサーバ：グループやメッセージの管理を行なうアプリケーションサーバ、連続音声認識コンソシアム (CSRC) のシステムをベースにした音声認識サーバ [5]、各種データが格納されるデータベースサーバなどから構成される。

5. 各クライアントの実装

各クライアントの機能（表 1）の詳細について述べる。

5.1 PC 環境

ウェブブラウザでユーザ認証およびグループ選択を行うと、メディアプレイヤーを模して設計された VoiceCafe ブラウザ（図 1）が起動する。ボタンを「録音」「戻る」「再生」「進む」「停止」「取消」「送信」「更新」に限定し、スレッド単位でメッセージを連続再生できるようにして操作を効率化した。ディクテーション結果のツリー表示（画面左部）およびスレッド一覧（画面右部）をクリックすることで当該箇所へのジャンプができる。

クライアント OS として Microsoft Windows を対象としているが、実装には Java (J2SE) および Swing, Java Media Framework, Java Web Start などプラットフォーム非依存の技術を使用した。

[†] 東京大学, University of Tokyo

[‡] (株) ネットイン京都, NetinKyoto Corporation

[§] 特定非営利法人 SCCJ, SCCJ

表 1: 各クライアントの機能

クライアント	PC	電話	PDA
ユーザ認証	e-mail	ID 番号	× *1
グループ操作	作成/選択	選択	× *2
グループ参加	承認/要求	要求	×
スレッド操作			× *3
再生			
発言			
全二重機能 *4			×
テキスト表示		×	×
テキスト修正	*5	×	×
操作手段	GUI	音声/DTMF	音声/ボタン
ガイド音声	× *6	*7	*8
その他の機能	*9	×	*10

- *1 ユーザ認証には無線インターネット (MISP) の認証を利用。
- *2 各ユーザのボイスメモ (留守電) を自動的に選択。
- *3 スレッド管理 (メッセージ関連付け) は未実装。
- *4 登録された発言を聞きながら返答を録音できる。
- *5 ディクテーションの誤りを文字入力により修正する機能。
- *6 Java (Swing) に対応したスクリーンリーダーの併用を想定。
- *7 録音音声および合成音声によるプロンプト。
- *8 合成音声によるメニューなどの読み上げ機能。
- *9 ウェブ連携 (メッセージに URL を関連づけできる)。
- *10 電話 (NOTASIP), ニュースリーダー。

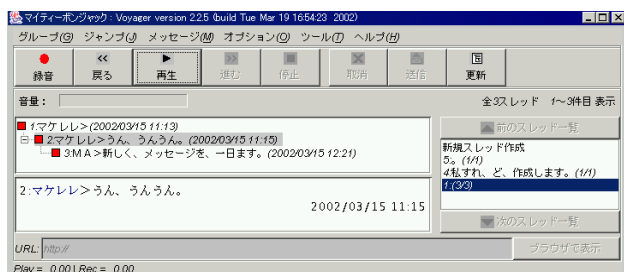


図 1: PC 用 VoiceCafe ブラウザの画面

5.2 電話環境

電話回線からの操作に対応して HTTP 経由でウェブサーバに要求を送るためのゲートウェイ設備 (システム構成上はクライアントの 1 つ) を構築した。グループを選択した後のメッセージ操作は、視覚障害者用スクリーンリーダーで音声化された Windows エクスプローラをカーソルキーで操作する場合のインタフェースを模している。

音声 (孤立単語音声認識) または数字キー (DTMF) により「次のスレッド」「前のスレッド」「次のメッセージ」「前のメッセージ」「新規の録音」「返信の録音」などのコマンドが使用できる。操作へのフィードバックとして移動先のメッセージが自動的に再生される。実装は Windows 上で C++ 言語によって行ない、電話音声認識サーバとして Nuance システムを、電話回線 I/F として Dialogic 社のボードを使用した。

5.3 PDA 環境

我々は視覚に頼らず操作できる PDA の開発を行っており、公衆無線インターネット「みあこネット」[6] 対応ドライバ、NOTASIP [7] 仕様のインターネット電話、ウェブの情報をダウンロードして合成音声で読み上げるニュースリーダーなどを Microsoft PocketPC 2002 上に実装している。この PDA において、VoiceCafe をボイスメモまたは留守番電話として利用するための機能を実現した。操作には上下ボタンと「決定」「取消」の 4 つのボタンを使用し、操作へのフィードバックとして合成音声によるメニュー項目の読み上げを行なう。実装には商

用の PDA 向け音声合成・音声認識エンジンを用いた。

6. 評価と考察

第 3 章で述べた要求に関する評価および考察を行なう。要求 (1) について、登録済みのメッセージに対して被験者 9 人が 12 回ずつ全二重機能を用いた返答操作を行ない、メッセージ関連付け誤差および相槌判定率を求めた (継続長 1.0 秒未満の発言を相槌と見なした)。PC および電話の両環境における各結果 (表 2, 表 3) は良好であり、発言の継続時間やタイミングを活かした操作を実現できた。

表 2: メッセージ関連付け誤差の評価結果 (平均)

クライアント	発言	誤差 (秒)
PC	相槌	0.31
	非相槌	0.35
電話	相槌	0.27
	非相槌	0.33

表 3: 相槌判定率の評価結果 (平均)

クライアント	発言	判定率 (%)
PC	相槌	96
	非相槌	96
電話	相槌	100
	非相槌	98

要求 (2) については、PC 環境のみを想定した AVM の設計 [2] (クライアント/サーバの 2 階層) を 3 階層化することにより、電話環境や PDA 環境への拡張はバックエンドサーバの変更を最小限にしつつ実現された。

要求 (3) については、PC 環境 ([2] での評価) に加えて、電話 / PDA 環境においては視覚障害者に馴染みやすく使いやすいインタフェースを実現できた。現在は視覚障害者による評価実験を準備している。

7. まとめ

PC, 電話, PDA から利用できる非同期型音声会議システム VoiceCafe の設計および実装について論じた。今後はネットワーク環境に固有の問題 (アクセス集中時の応答遅延や通信回線が不安定な場合など) に柔軟に対処できることを検証したい。また、ディクテーション性能、導入や管理のしやすさ、使いやすさなどの面から、システムの拡張と改良を行いたい。

謝辞

開発に参加・協力いただいた多くの方々へ感謝する。なお、本システムの開発においては、科学技術振興事業団・平成 11 年度委託開発課題「コミュニティ形成支援機能を持つ音声会議システム」および経済産業省・平成 14 年度 e! プロジェクト委託事業「地域情報基盤におけるコンテンツ配信とピアツーピア環境の構築」の支援を受けた。

参考文献

- [1] 西本卓也 他: 非同期音声メッセージシステムの設計, 信学技報, MVE97-98, pp.39-46, Feb. 1998.
- [2] 西本卓也 他: 非同期型音声会議システム AVM の設計と評価, 信学論 D-II, Vol.J83-D-II, No.11, pp.2490-2497, 2000.
- [3] www.voicecafe.jp, www.dsneo.co.jp
- [4] 例えば www.blogger.com, www.audioblogger.com など.
- [5] 川原毅彦 他: 非同期型音声会議におけるディクテーション機能, 日本音響学会 2001 年春季講演論文集, 1-3-20, 2001-03.
- [6] www.miako.net
- [7] www.notasip.org