

オンラインゲームにおける匿名性を有した音声チャットの開発

1. 背景

オンラインシミュレーションゲームでのコミュニケーションの手段は、文字チャットが一般的であるため、タイピングが遅いという理由だけでこのようなゲームを十分に楽しめないユーザも多い。かねてから、オンラインシミュレーションゲームにおいて、音声を利用したチャット機能の搭載が望まれている。一方で、音声チャットが実装されたとしても、音声には個人を特定する情報が含まれているため、不特定多数のユーザを相手にするオンラインゲームでは利用を好まないユーザもいる。

2. 目的

上記の状況を踏まえ、本プロジェクトでは、音声に含まれる個人を特定する情報を削除することにより匿名性を保ちながらも、音声に含まれる感情情報を伝達することで表情豊かなコミュニケーションを実現する音声チャットを開発する。特にオンラインゲームのような不特定多数のユーザとのカジュアルなコミュニケーションを楽しむ場においては、安心して会話を楽しむことができる点で有用であると考えられる。

3. 開発の内容

3. 1. システムの構成

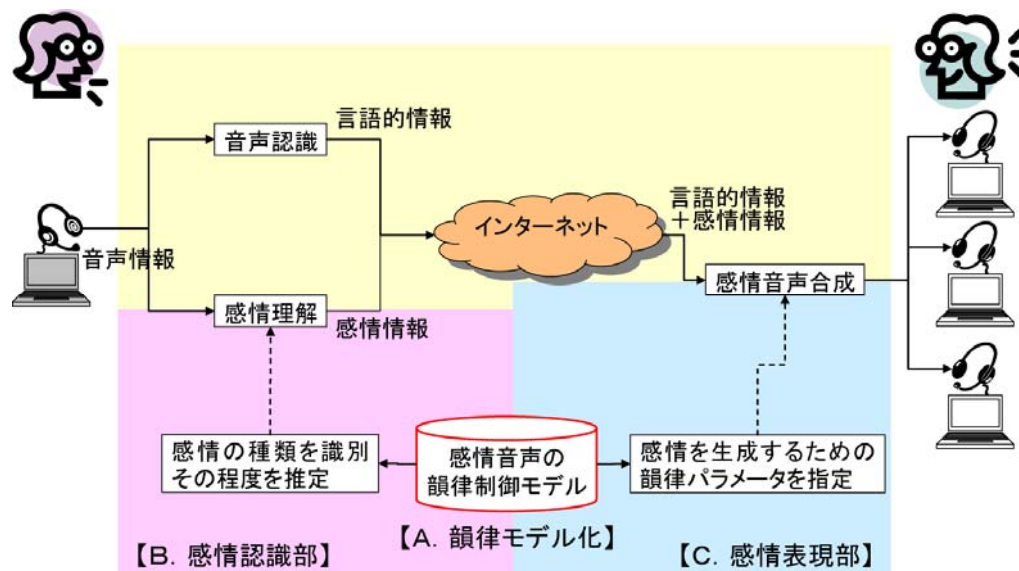


図 1. 匿名性を有した音声チャットのシステム構成図

本プロジェクトで開発した匿名性を有する音声チャットのシステム構成図を図1に示す。音声チャットは、オンラインゲームへの組み込みは行わず、様々なソフトで利用できるように独立したアプリケーションとした。言語的情報と感情情報とをインターネットを介して伝達するためにスカイプのテキストチャットの機能を利用した。

スカイプは、ファイアウォールや NAT の内側にあるパソコンからも、特別な設定を行なうことなく接続できる特徴を持つ。また、スカイプの機能を拡張・利用するための API が公開されている。

3. 2. 感情認識部

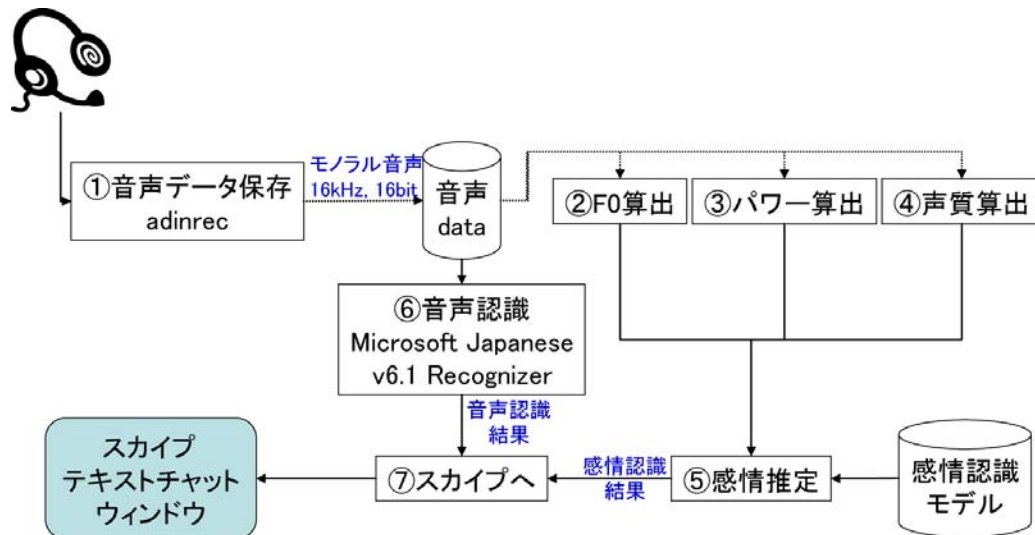


図 2. 音声認識，感情認識のブロック図

図 2 に感情認識部のブロック図を示す。ユーザの発した音声は、発話単位に切り出され、音声データとして保存される。保存した音声データを対象に、音声特徴量の算出を行ない、構築した感情認識モデルを用いて感情の推定を行なう。同時に、音声認識エンジンを利用して音声認識を行なう。音声認識，感情認識の結果は、スカイプのチャットメッセージとして入力される。以下にそれぞれの機能について説明する。

①音声データ保存

音声切り出し録音ツールである adinrec を用いて、マイク入力の音声区間の検出を行ないながら、音声を保存する。波形の振幅レベルのしきい値は 3000 とする。

②F0 算出

音声分析・合成ツールである Praat を利用し、10ms 間隔で F0 値の抽出を行なう。抽出した F0 パターンから F0 値の最高値，最低値，標準偏差を求める。

③パワー算出

F0 値と同様に Praat を用いてパワーの抽出を行なった。Praat では音声の振幅情報をもとにパワーの抽出を行なうが、より正確な抽出を行なうために F0 の情報と同期した抽出を行なっている。抽出したパワーの情報に基づきパワーの最大値，標準偏差，変動量を求める。

④声質算出

声質に関する特徴量であるケプストラム係数のうち，その第 1 次係数を抽出し，平均値と標準偏差を求める。ケプストラム係数の抽出には Praat を利用している。

⑤感情推定

感情推定のための音声特徴量を受け取り、この音声特徴量に対し、構築した感情認識モデルに基づき、感情の種類とその程度の推定を行なう。

⑥音声認識

発話単位で保存されたユーザの音声を、音声認識することによりテキスト化する。音声認識は、フリーの音声認識エンジンである Microsoft Japanese v6.1 Recognizer を利用している。音声認識エンジンの機能を開発システムに組み込むために、Speech API 5.1 を利用する。

⑦スカイプへ

音声認識と感情認識の結果を受け取り、この結果をスカイプのチャットメッセージとして入力する。Skype API を利用する際、COM コンポーネントとして Skype4COM を利用している。

3. 3. 感情表現部

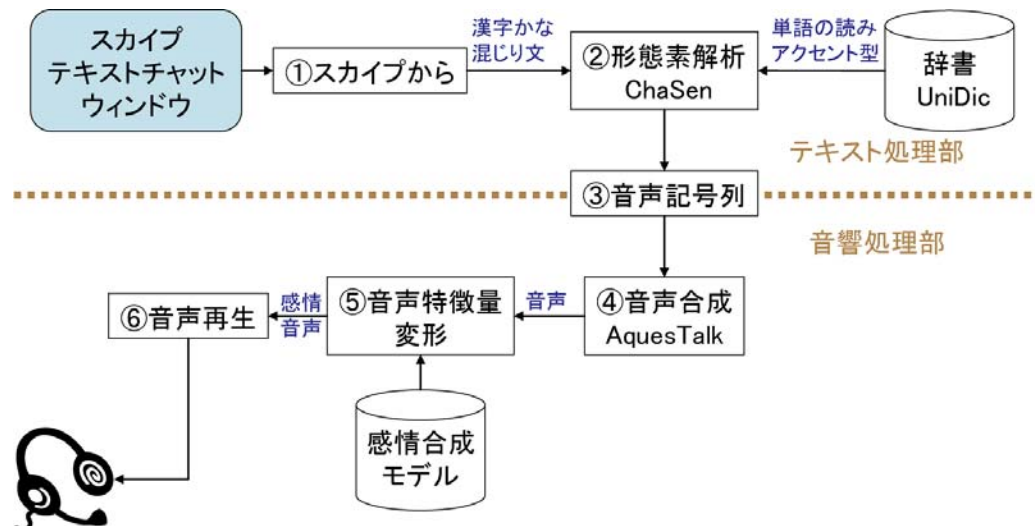


図 3. 感情音声合成のブロック図

スカイプのテキストチャットで受信したメッセージの音声化を行なう。図3に感情音声合成のブロック図を示す。受信メッセージに対し、形態素解析などのテキスト処理を行ない、読み上げに必要な情報を付与する。この情報に基づき、フリーの音声合成エンジンを利用し、平静音声を作成する。作成した平静音声に対して、構築した感情合成モデルに従い、音響的な変形を加えて感情音声を作成する。

①スカイプから

Skype4COM を利用して、テキストチャットのメッセージを受け取る。

②形態素解析

受け取った漢字かな混じりのメッセージを形態素解析器 ChaSen により形態素単位に分割する。解析の際、茶釜用形態素解析辞書 (UniDic-chasen-1.3.5) を用いて、形態素に対し、読みとアクセント型を付与する。さらに、形態素を文節単位に結合

して、その文節に対してアクセント結合規則を用いてアクセント型を付与する。

③音声記号列

文節の読みの情報とアクセント型の情報に基づき、音声合成エンジンの入力フォーマットに従った音声記号列化処理を行なう。

④音声合成 AquesTalk

音声記号列に従い、音声を合成する。合成時の声の種類として、女声 2 種、男声 2 種、ロボット声 1 種、中性的な声 1 種、機械的な声 2 種が用意されている。

⑤音声特微量変形

AquesTalk が合成した平静音声の音声特微量に対して、構築した感情合成モデルに基づき韻律情報の変形を行ない、変形後の韻律情報に従い感情音声を合成する。音声特微量の抽出と合成には、Praat を利用する。

⑥音声再生

Praat を利用し、合成した感情音声を再生する。

4. 従来の技術（または機能）との相違

音声の匿名性を確保する手段として、声質変換の技術があるが、従来の声質変換では、音声の自然性を保ったまま韻律情報を変換するのは困難である。そこで本プロジェクトでは、音声認識と音声合成の技術を利用し匿名性を確保するという新しいアイデアに基づいて開発を行なった。ゲーム中、表情豊かなコミュニケーションを実現するために感情情報に着目した。感情情報を伝達する際に必要となる感情認識、感情音声合成のためのモデルは、それぞれ高精度なものを構築することができた。

5. 期待される効果

合成時に基本となる音声を 8 種類用意しており、ユーザはいくつかの特徴をもつ音声の中から、アバターの音声を選択することが可能となり、オンラインゲームをする上での新たな面白さとなる。また本システムでは音声を認識して、言語的な情報として扱うため、翻訳機の利用が容易であると考えている。翻訳機を利用すれば、言語の異なるユーザ同士のコミュニケーションが可能となる。さらには、音声により推定したユーザの感情は、表情などを変化させることによりアバターに反映することが可能であり、シミュレーションゲームでのリアリティの向上につながると考えている。

6. 普及（または活用）の見通し

オンラインゲームへの登録ユーザ数は年々増え続けている。匿名性が確保されないことを理由に、オンラインゲームにおいて音声チャットを利用していなかった多くのユーザに、本プロジェクトで開発したシステムを利用していただきたいと思っている。

7. 開発者名（所属）

河津宏美（東京工科大学大学院）