

タイトル: ユーザからの感情値によって音声感情と言語の反応を変化させるボイスアシスタント

1. はじめに

近年、音声認識技術や自然言語処理の発展に伴い、音声対話システムを応用したボイスアシスタントシステムが普及してきている。ボイスアシスタントシステムをはじめとした音声対話システムは従来のCUI (Character User Interface)やGUI (Graphical User Interface)で構築されたシステムとは異なり、人対人のコミュニケーションのようにコンピュータを制御することができる。そのため、年々機能が増え、複雑化するコンピュータを使用するのに適したインタフェースになりうるかと期待されている[1]。例えば、AppleのiOSに搭載されているボイスアシスタントである「Siri」はアラームの設定やWeb検索だけでなく、ユーザがスマートフォンに登録している連絡先などを参照し、電話を発信したり、メッセージを送信することができる。

しかし、このようなシステムは利用率が低いという現状がある。この原因を調査し、「声のリアリティ」や「返答が単調でない」などのポジティブな類人間性を実装し、従来のシステムと検証した研究[2]がある。これに対し、本研究ではシステム内の仮想エージェントに喜びなどのポジティブな感情だけでなく、怒りなどのネガティブな感情表現を実装し、検証を行った。

2. 提案システム

本研究では「中立、喜び、怒り」の感情をユーザの音声感情に合わせて、エージェントが表現する音声アシスタントシステムをPythonを用いて制作した。システムのフローチャートを図1に示す。本システムはテキスト処理部、感情分析部、出力部に分けられる。音声認識部ではユーザがEnterキーを入力した直後から5秒の間、マイク入力された音声を録音する。その音声をGoogle Speech Recognition APIに送信し、返答された結果を音声認識されたテキストとして処理した。また、タスク-雑談識別機では、音声認識されたテキストからユーザがシステムに対し雑談を望んでいるのか、またはタスクの実行を望んでいるのかをGPT APIを用いて識別した。

感情分析部では、音声認識部で録音した音声をEmpath APIへ送信し、その返答を処理した。Empath APIは送信した音声に対して、「平常、怒り、喜び、悲しみ、元気度」の5つの感情にそれぞれ当てはまる度合を0以上50以下で出力する。本システムでは「平常、怒り、喜び」の値のみを参照し、最も値が高い感情をユーザが入力した音声感情として処理した。

最後の出力部では、感情分析部と雑談-タスク識別機から出力された分析結果の組み合わせ、計6通りのプロンプトとGPT APIを用いてユーザに提示する出力テキストを生成した。雑談-タスク識別機でタスクと認識された場合のプロンプトには実行可能なタスク(メモ帳の起動、ブラウザ検索)を実行するためのPythonコードを状況に応じて出力テキストと共に出力するように調整した。また、出力テキストからVOICEVOXを用いて音声合成し、出力テキストと同時に提示した。

