

C2 音声インターフェースによるインスタントメッセージングエージェントの開発

窪田 元

1. はじめに

近年、コンピュータによる情報の交換は大規模化し、より高い即応性が求められてきている。ユーザ層も多様化し、手が不自由な人や、キーボードを全く触ったことのない人たちとの格差も深刻である。そこで、音声インターフェースを用いたインスタントメッセージングエージェントシステムを提案する。よりスムーズ、かつ誰にとっても簡単にチャットを行う環境を実現するためである。

2. 要素技術

2.1 Speech SDK[1]

音声認識や音声合成を使うための Windows 対応の SDK のことである。認識、文法構成や合成などはランタイムコンポーネントの sapi.dll で行われる。この SDK は C++ と VC のどちらかの言語で開発可能であるものであり、本件急で用いる Java で開発する MSN Messenger API[3]に対応するため JNI[4]を用いる。日本語音声合成は、ドキュメントトーカ[2]を用いる。

2.2 インスタントメッセージングサービス

インターネットを通じてユーザ同士でリアルタイムにメッセージを交換するシステムである。本システムでは MSN メッセンジャーを用いる。[5]によるシステムへ加える、新たなる付属機能として実現する。

3. メッセンジャーエージェント

3.1 システム構成

本システムは主に音声認識、音声合成の機能が中心となっている。MSN Messenger API (msnAPI) と Speech SDK に含まれる Speech API (SAPI) を利用し、音声認識、音声合成を可能にしている。本システムのシステム構成図を図1に示す。

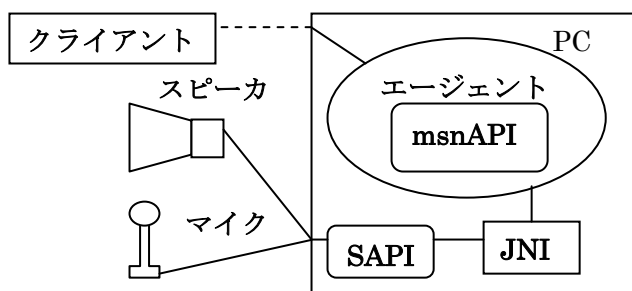


図1. システム構成図

音声認識では、マイクから習得した音声データを SAPI により音声認識し文字化、JNI[4]を用いてエージェントへとデータを受け渡す。ここでその入力された文字のコマンドに準じた機能を、msnAPI を使用し呼び起こす。ここからクライアントとのやり取りが始まる。

音声合成ではエージェントがクライアントから文字を受信、JNI を使用しデータを受け渡し、SAPI とドキュメントトーカによりその文字を音声化、スピーカで発声される。

4. 実験

本システムの有効性を評価するために、チャット会議機能を用いた次の実験を行って音声入力、キーボード入力との差を比べた。2008年12月12日に、2人1組で合計100単語分のしりとりを行った。片方は、音声入力によりメッセージを送信し音声合成による受信を行う。尚、しりとりにおける思考時間は考慮しない。図2に音声入力とキーボード入力の入力時間を示す。

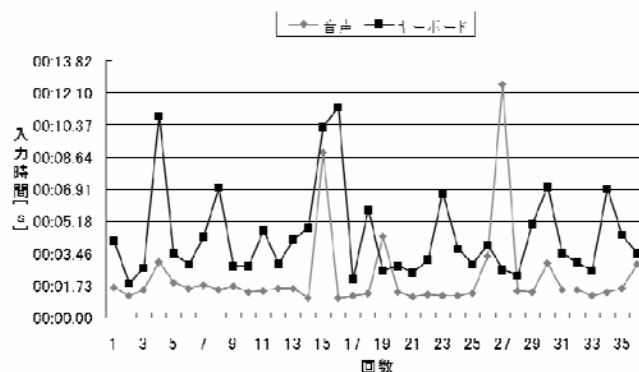


図2. インターフェースによる入力時間分布

音声入力では15、27回目など、異常に時間がかかっている時がある。これは誤認識が原因である。キーボード入力においても、4、15回目で時間がかかっているが、変換ミスなどが原因である。表1にそれぞれの誤認識回数及び平均入力時間を示す。同じ実験を合計3名で行ない、同様の結果となることを確認した。

表1. 入力誤認識及び平均

	音声	キーボード
誤認識[回/35 回中]	15	0
平均入力時間[s]	2.18	4.39

5. おわりに

メッセンジャーエージェントに音声認識機能、音声合成機能を追加し、より利便性を向上させたシステムを開発し、その性能を評価した。今後の課題として、音声認識の認識精度の問題がある。

参考文献

- [1]Speech SDK
http://www1.bbiq.jp/kougaku/speech_reco.html (2008 年 12 月参照)
- [2]ドキュメントトーカ 日本語音声合成エンジン for Windows
<http://www.createsystem.co.jp/DTalkerSapi1.html> (2008 年 12 月参照)
- [3]MSN Messenger API
<http://jmsn.sourceforge.net> (2008 年 12 月参照)
- [4]JNI:Java Native Interface(2008 年 12 月参照)
<http://www.alles.or.jp/~torutk/ojava/maneuver/2001/jni/jni.html>
- [5]佐上, 陳, インスタントメッセージによるお天気配信サービス “Kikn エージェント君”, 東海大学電子情報学部, 2007 卒業論文, 2008.