

デー3 インターネットフロー情報を対象にした 音声インタフェースエージェント

NEPIA (NEws Program with Interface Agents) *

金子和恵 廣田誠 八木沢津義

キヤノン (株) プラットフォーム技術開発センター

1 はじめに

音声合成・音声認識を用いた新しい Human Computer Interaction 技術の一つとして、インタフェースエージェント技術が注目されている。これは、CG やアニメーションの技術と音声合成・音声認識技術を組み合わせ、コンピュータ上に作り上げたキャラクターと音声を介してコミュニケーションするものである。

このインタフェースエージェントの可能性を確かめるために、キヤノンの音声合成エンジン [1] と音声認識エンジン [2] に、キャラクターアニメーション技術を組み合わせたインタラクティブなニュース番組のデモを作成した。これは、8名のキャラクターが、ユーザの音声による指示に応じて、それぞれの合成音で、インターネット上のニュース (<http://www.asahi.com>) を読み上げるものである。

2 インタフェースエージェント

インタフェースエージェントの枠組みとして、Microsoft Agent [3] を採用した。この枠組みでは音声合成エンジンや音声認識エンジンの選択が自由である。これに、キヤノンの音声合成エンジンと音声認識エンジンを組み合わせた。

キャラクターとしては8名 (図1参照) を設定し、年齢や性別のバリエーションを持たせた。個々のキャラクターには、年齢・性別にあわせた合成音声を割り当てており、女の子のキャラクターは女の子らしい声で、年配の男性のキャラクターは年配の男性らしい声で発声する。

また、キャラクターのアニメーションの種類は、表情 (驚き、恐怖、嫌悪、怒り、幸福、悲しみ、ニュートラル) を持って発話内容に同期した唇の動きをするものの

他に、相づちや指し示しなどの25種類の動作などを用意しており、感情豊かな表現を行う。

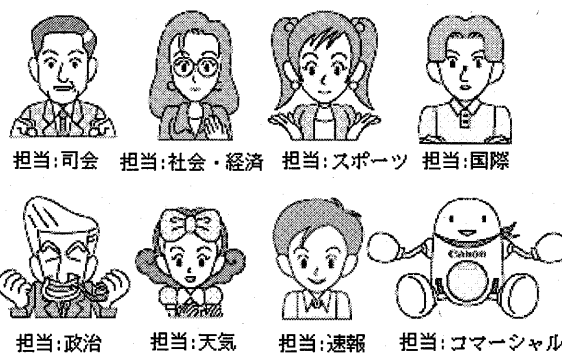


図1. キャラクター

3 ニュース記事の参照

ニュース収集プログラムは、起動時に指定ホームページを参照し、そのHTMLの構造を解析してリンクをたどってコーナー毎にニュース記事を振り分ける。このデモでは、対象を <http://www.asahi.com> に限定している。

また一定時間ごとにホームページを参照して、更新内容があれば、速報記事として正規のコーナーとコーナーの間にさしはさむというを行っており、刻一刻と変化する情報を、ニュース速報として割り込み提示するという試みも行っている。

4 ニュース番組の構成

番組は、社会、経済、スポーツ、政治などのコーナーに分かれ、インタフェースエージェントの個々のキャラクターがそれぞれのコーナー (図1参照) を担当する。ニュース番組をスタートすると、社会、経済、スポーツ、政治といったあらかじめ設定したコーナーの順番で、キャラクターがニュースを読み上げて行く。この使い方では、普通のニュース番組と同じであるが、ユーザは途中で自分の聞きたいコーナーを選択できる。現在アナウンス中のキャラクターに対し、自分の聞きたい

*NEPIA (NEws Program with Interface Agents): The Interface Agents that announce flow information on the Internet

KANEKO Kazue, HIROTA Makoto and YAGISAWA tsuyoshi (Platform Technology Development Center, Canon Inc.)

コーナーを音声で指示すれば、そのコーナーを担当するキャラクターに画面を譲って番組を続ける。

ニュース番組のそれぞれのコーナーでのキャラクターの発話内容や動作は、独自の簡単なシナリオ記述言語で記述している。コーナーが指定されると、そのニュース記事からシナリオを作成し、シナリオ通りの動作を進めて行く。

5 実行例

NEPIA の実行画面を図2に示す。テレビ画面をイメージしたシンプルなもの、右側に認識可能な文のセットを表示している。左側の画面上で、キャラクターがアナウンサーとしてニュースを読み上げている。

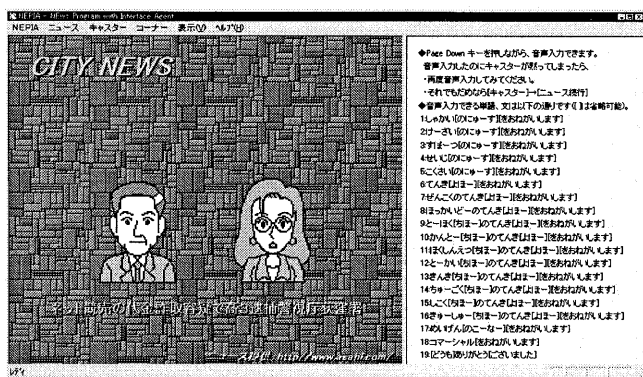


図2. 実行画面1

ニュース内容とそれを読むキャラクターの感情表現を同調させる試みの第一歩として、お天気のコーナーでは、今後3日間の天気が晴れがちならば笑顔(図3参照)、曇りがちならば嫌悪、雨がちならば悲しみ、といった表情で天気予報をアナウンスするようにしている。

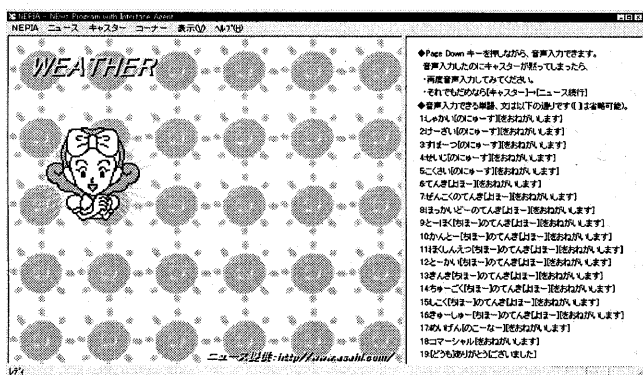


図3. 実行画面2

なお、キャラクターの感情表現を有効に利用するために、コーナーの合間に寸劇風のコマーシャルを差し挟んでいる。これは、独自のシナリオ記述言語で予め記述したもので、動的に生成したものではない。

6 まとめ

インタフェースエージェントの可能性を確かめるためにニュース番組のデモを作成した。日本語解析と合成音声の品質から、聞き流しながらもニュースの概要をある程度把握できるという印象を得ている。集中しなくては把握できない文字情報を耳で聞くことによって、インターネット上のフロー情報を「ながら」的に取りこむという利用形態も可能になるものと思われる。また、アニメーションの親しみやすさから気楽に楽しめるニュース番組にもなっている。

今後の課題としては、以下のことがあげられる。

- 目で読んで分かりやすい記事でも、耳で聞くと聞き取りにくい場合がある。略語や体言止めなどを、聞き取りやすい語や言い回しに変換するなどの言語処理が必要である。
- 動作や表情を豊かにすると、合成音声は棒読みに聞こえる。動作や表情によって、合成音声に感情を含んでいるように錯覚する場合もあるが、自然な音声に聞こえるには、抑揚や感情を含んだ合成音声が必要である。
- 聞きたいコーナーの指定はできるが、聞きたいニュースの指定はできない。ユーザの選択できる幅が狭いため、インタラクティブという点では弱い。ニュース記事を分析してユーザが興味を持っているトピックを指定できるようにするなど、検索機能や動的な認識語彙の切り換え機能などを取りこむ必要がある。

【謝辞】

今研究に対し、<http://www.asahi.com> 上のニュースを提供し協力してくださった朝日新聞社に心から感謝致します。

参考文献

- [1] Yamada et al.: PureTalk: A High Quality Japanese Text-to-speech System, (International Conference on Spoken Language Processing), ICSLP 2000 (掲載予定)
- [2] 山田他 パラメータのスカラ量子化と混合分布 HMM の次元独立演算による高速出力確率計算, 信学技報, SP-95-222, 1995.
- [3] Microsoft Corporation: Microsoft Agent プログラミング技法, 日経 P B ソフトプレス, 1998.