



平成11年9月10日

日本電信電話株式会社

キャラクタが感情をこめて合成音声で話す『WebMessenger』を開発
—親しみやすく訴求力のあるマルチメディアコンテンツの作成を支援—

NTTは、口調を自由に換えられる合成音声とCGアニメを組み合わせ、感情のこもった表現を可能にする新しいインタフェース技術

『WebMessenger』を開発しました。『WebMessenger』は、WWWブラウザに組み込んで使えるように設計されています。

インターネットが普及するにつれ、様々なホームページが開設され、誰でもコンテンツの作成や情報発信がおこなえる時代に入りました。しかし、ホームページ等における表現に関しては、テキストが主体で、音声や動画を含むコンテンツを扱うとデータ送信や表示に時間を要するため、魅力的なマルチメディアコンテンツ作成に限界がありました。

『WebMessenger』では、文意に応じて口調を自在に編集して感情をこめた合成音声を作成できます。さらに、合成音声と50～60パターン用意されたCGアニメ動作表現を自然なタイミングで連動させることが可能です。また、合成音声単位データ、CGアニメデータを予め情報の受発信側で共有し、データを選択するインデックス情報のみを伝送する技術を採用しています。このため、少ない伝送データ量で、合成音声と同期したCGアニメの動きを楽しむことができます。

『WebMessenger』は、NTTサイバースペース研究所が開発した『音声合成システム』（テキスト情報から音声合成を行うシステム）をベースに、『Sesign98』（合成した音声情報の加工を行う合成音声デザインツール）を要素技術として、合成音声とCGアニメとを自由なタイミングで同期させる仕組みを加えて開発されました。

主な用途には、遠隔教育システム等での子供向けの親しみやすいインタフェースやホームページの内容を分りやすく読み上げてくれる音声エージェン

トなど、マルチメディアコンテンツ作成および閲覧用のツールとして広範囲な利用が考えられます。

なお、『WebMessenger』を使用したデモシステムを、9月14日～16日に開催されるICCC'99 EXPO(*1)に展示します。

<システム構成> ([システム構成図参照](#))

本技術では、合成音声単位データ、CGアニメデータを予め情報の受発信側で共有します。テキストから合成音声へ変換し、合成音声のイントネーションなどの加工は「Sesign98」を用いて、自動変換（お任せモード）のほか、制作者の好みにより任意の口調や速度に設定することができます。こうして作成した合成音声データを「WebMessenger-Creator」により50～60パターンの感情表現を持つCGアニメと自由に組み合わせることが可能です。

作成された合成音声、CGアニメはそれぞれ、“合成単位インデックス”、“動画インデックス”と呼ばれるインデックス情報により表現されます。情報発信者は、このインデックス情報のみをHTML文書に添付してインターネットに送信するので、合成音声やCGアニメを少ない情報量で伝えることができます。閲覧側では、このインデックス情報をもとに『WebMessenger-Player』が、合成音声とCGアニメの単位データをインデックス情報に従い結合し、合成音声とCGアニメを再生します。

<主な特徴>

1) 音声の編集機能により感情のこもった表現が可能

ワープロソフトでフォントの種類や色、大きさを設定するように、それぞれの合成音声について音質・発声速度・抑揚などが自由にデザインできます。例えば、ささやくような「おはよう」と、元気いっぱいのおっはよ〜！を使い分けることもできます。

2) CGアニメの動きと合成音声がぴったり同期

従来の同種ソフトでは、アニメの動きと音声が必ずしも一致していませんでした。本技術では、両者の同期はもとより、例えば、「挨拶」の動作途中から「こんにちは」としゃべらせる、といった細かいタイミング設定も可能です。

3) アニメと言葉の対応づけも自由自在

時として言葉とは裏腹のしぐさをするのが人間です。

『WebMessenger』では、「頑張ります」という言葉に、「しぶしぶ」のしぐさをしたアニメを対応させるなど、言葉と動作を自由に組み合わせることができるので、より人間っぽい表現、など、バラエティ豊かなコンテンツ上の表現が広がります。

4)わずかな伝送量で映像伝送と同じ効果

インターネット上で送信されるのは合成音声単位インデックスと動画インデックス情報のみのため、同じ映像効果を得るために必要な映像伝送と比べて、データ量が約1/120となります。

<技術のポイント>

1)合成音声とCGアニメの同期調整技術

『WebMessenger』では、合成音声とCGアニメの情報を、音韻情報・イントネーション等の音声の詳細な情報に、CGアニメーションの動作情報を含める形式で記述しています。これにより、合成音声をCGアニメと任意のタイミングで再生することを可能としました。

2)Sesign98による音声の自在な編集

合成音声編集は、NTTが独自に開発した音声デザインツール「Sesign98」を使用します。Sesign98は、GUIにより簡単に合成音声の大きさ、高さ、声質などを変更できます。また、一度作成した合成音声をライブラリー化することもできるため、誰でも手軽に合成音声を作成および編集することが可能です。

<主な用途>

自由度の高い設計となっているため、様々な用途が考えられますが、コンテンツの送り手と受け手が、同じプログラム、データ（音声の合成単位データとCGアニメデータ）を共有するシステムのため、会員制システムなど、閉じたユーザー間での利用には最も適しているといえます。ここでは、教育現場での応用例をふたつ提案します。

1)統合教育システム「CALAT」

CAIシステムのユーザインタフェースとして、『WebMessenger』を用いれば効果的です。たとえば、NTTの開発した統合学習システム

CALAT(Computer Aided Learning and Authoring Environment for Tele-Education)では、サーバ上で稼働している独自の学習支援プログラムが、個々の学習者の学習状況や理解度に応じて学習内容や進み具合を調整するなどひとりひとりに見合った学習環境を提供しますが、このインタフェースとして用いて、CGアニメによる温かい励ましや評価の表現によって、学習者の意欲をさらに高めることが可能です。

2) 「サイバー学級日誌」

テキストからの音声合成が効果を発揮する例としては、学級日誌をホームページ上で公開する「サイバー学級日誌」なども考えられます。任意のテキストを音声に変換し、感情を付与することもできるので、文章だけの日誌とはひと味違う日誌づくりが可能です。もちろん、動物の観察記録や社会見学などの様々な体験に感情を込めて報告や伝達する用途にも応用できます。

* 1 : ICCC'99 EXPO

国際コンピュータ通信会議(International Conference on Computer Communication)は、ICCC協議会(International Council for Computer Communication)をスポンサーに、通信系事業者が中心となって、1972年以降隔年に開催されてきた。今年、1978年以来21年ぶりに日本で開催。

「ICCC '99 EXPO」は、その併催展で「コンピュータ・通信・放送・家電のデジタル統合にもとづく多様な発展」をテーマに開催される。

会 期：1999年9月14日(火)～9月16日(木)の3日間 10:00-17:00

会 場：東京国際フォーラム 展示ホール

(詳細は、http://www.convention.co.jp/iccc_j/を参照下さい)

別紙

- ・ [システム構成図](#)

<本件に関する問い合わせ先>

NTTサイバーコミュニケーション総合研究所 情報戦略担当 川嶋・萩野
TEL : 0468-59-2032

e-mail : ckoho@tamail.rdc.ntt.co.jp



[NTT NEWS RELEASE](#)