

iFACe : インタラクティブ発話合成システム

四 倉 達 夫[†] 川 本 真 一[†]
松 田 繁 樹[†] 中 村 哲[†]

iFACe: Interactive Facial Animation Control system

TATSUO YOTSUKURA,[†] SHINICHI KAWAMOTO,[†] SHIGEKI MATSUDA[†]
and SATOSHI NAKAMURA[†]

1. はじめに

本稿では誰でもが声優を体験可能なインタラクティブ発話合成システム : iFACe を提案する。現在の日本アニメの制作において、キャラクターの音声を吹き込む場合、一般的に「アフレコ」を行う。アフレコは、あらかじめ作成されたキャラクターの口の動きに、声を合わせる必要があり、プロの声優が行う専門作業である。iFACeはブレンドシェーブ・キーフレームベースのリップシンクアニメーションシステムで、ユーザが発話した任意の音声を入力とする。アミューズメント施設のような雑音環境下においても、システムは音声から音素・音素継続長を正確に推定し、その結果をもとに Cartoon CG キャラクタにあった口形状のキーの自動配置を行う。さらに出力されたキーから Cartoon CG ヘマッピングを GPU 上で行い、リアルタイム発話アニメーションが生成可能となる。

iFACe のシステム評価として日本科学未来館にて 5 日間、親子連れや子どもをターゲットに展示を行った。システム動作回数は 1500 回を超えたが、システムは雑音環境下においても安定した出力結果を得ることができた。幅広い年齢層の参加者が iFACe に触れ、楽しんでもらうことができた。また期間中アンケート調査も実施し、システムのおもしろさ、アミューズメント施設へ応用化に関して高評価を得ることができ、エンターテインメント分野における iFACe の有効性を確認することができた。

2. 特 徴

本システムの関連研究として Better Face Communication System[1]、Galatea FSM[2] があるが、本システムの特徴は以下の通りである。

2.1 雑音環境下で動作可能な音声認識システム

エンターテインメント施設等、騒音の大きい場所で発話された音声でも高精度に音素セグメンテーションを行うため、当研究所で開発した音声認識ソフトウェア ATRASR [3] を使用した。マイクから入力された音声は、パーティクルフィルタを用いた背景雑音の除去手法 [4] により雑音抑圧される。また、子供用音響モデルをシステムに追加 [5] することで、大人用音響モデルだけでは認識しづらい子どもの声にも対応した。さらに、システムに登録していない台詞に対して、連続音素認識によるラベルなし音素セグメンテーションが行われる。

2.2 Cartoon キャラクタにマッチしたキーフレームアニメーション生成技術

2.1. で求めた分析結果の音素、音素継続長をキーフレームから Cartoon キャラクタに合ったキーフレームアニメーションの生成を行う。分析結果を直接キャラクタマッピングする手法 [1][2] を Cartoon キャラクタに適用すると、発話区間に多くのキーが存在するため、唇の動きが激しくなり「うるさいアニメーション」となる場合が多く、自然なアニメーションを実現する事が困難である。本システムでは、不要な音素キーを間引き、最適なキー配置にする手法 [6] を用いる、ことで Cartoon キャラクタでも自然な発話アニメーションの生成を可能とした。

[†] ATR 音声言語コミュニケーション研究所
ATR Spoken Language Communication Labs.

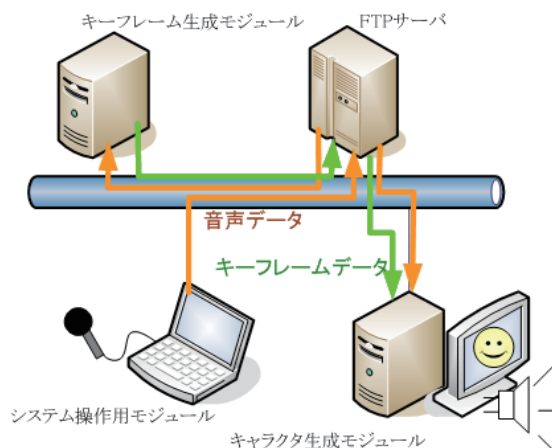


図1 システム概要
Fig. 1 System Overview

2.3 GPU ベースリアルタイムアニメーション

出力されたキーフレームから CG キャラクタへマッピングし、発話アニメーションを生成する。CG キャラクタは、3DCG ソフトウェアで作成された無表情および、「あ・い・う・え・お」の口形状を作成してあれば、トポロジの異なる複数の CG キャラクタをシステムへ登録することができる。また、複雑な顔形状をもつキャラクターに対し、リアルタイムアニメーションを実現するために、顔オブジェクトをグラフィック専用メモリに展開し、キーフレームアニメーションを GPU にて計算を行うよう設計されている。

3. システム概要

iFACe のシステム構成を図 1 に示す。モジュールは 1) システム操作モジュール、2) キーフレーム生成モジュール、3) キャラクタ生成モジュールで構成されている。ユーザはタッチパネルで発話したい台詞・もしくは自由発話を選択し、音声を入力する。入力された音声データはキーフレーム生成モジュールへ送られ、音声分析・キーフレーム生成を行う。キーフレームデータはキャラクター生成モジュールに送られ、ユーザが指定したキャラクターが収録した音声と同期しながら発話アニメーションを行う。データの受け渡しは FTP サーバを経由して行われる。

4. 評価実験およびまとめ

2006 年 5 月 3 日から 7 日まで日本科学未来館にて iFACe の展示を行い、延べ 1500 回以上の発話アニメーションを生成した。開催期間中、参加者に対してアンケートを配布し本システムの主観評価実験を実施した。有効回答数は 89 件であった。

評価実験は 5 段階評価実験方式で「非常に思う」を 5、「非常に思わない」を 1 とした。評価結果は以下のとおりである。

- エンターテインメントシステムとしての「おもしろさ」: 平均 4.62
- アミューズメントシステムやゲームで同様のシステムを遊んでみたいか: 平均 4.02
- 同様のシステムを携帯コンテンツとして遊んでみたいか: 平均 3.44

また自由記述形式での iFACe の意見では、有名なアニメ・俳優キャラクターで遊びたい、伝言・留守番電話サービスで使いたい、教育、医療現場で利用したいという意見が挙がった。参加者の中では日本語以外の外国語を、システムに入力させている場合もあったが、連続音素認識によるラベルなし音素セグメンテーション結果が優れ、概ね満足する生成結果を得ることができた。以上の評価結果からエンターテインメント分野における iFACe システムの有効性が確認できた。

謝辞 本研究の一部は JST, CREST 支援によるものである。

参考文献

- 1) Yotsukura, T., Fujii, E. and Morishima, S.: Face-to-Face Communication System in Cyberspace by Voice Driven Avatar, *EuroGraphics98 Short Presentation*, pp.2.6.1-2.6.4 (1998)
- 2) Yotsukura, T., Morishima, S. and Nakamura, S.: Model-based talking face synthesis for anthropomorphic spoken dialog agent system, *Proc. 11th ACM MM*, pp.351-354 (2003)
- 3) 伊藤 玄, 葦苜 豊, 實廣 貴敏, 中村 哲: 音声認識統合環境 ATRASR の概要と評価報告, 日本音響学会 2004 年秋期研究発表会講演論文集, 1-P-30 (2004)
- 4) Fujimoto, M. and Nakamura, S.: A Non-stationary Noise Suppression Method Based on Particle Filtering and Polyak Averaging, *IEICE Trans. on Information and Systems*, Vol.E89-D, No.3, pp.922-930 (2006)
- 5) 松田 繁樹, 實廣 貴敏, 中村 哲, 石井 カルロス 寿憲, 神田 崇行: コミュニケーションロボットにおける音声認識性能の評価, 日本音響学会 2005 年秋期研究発表会講演論文集, 2-P-22 (2005)
- 6) Kawamoto, S., Yotsukura, T. and Nakamura, S.: Key-frame Removal Method for Blendshape-based Cartoon Lip-sync Animation, *ACM SIGGRAPH 2006 Research Posters*, 0202-kawamoto.pdf (2006)