

# ビジュアルコミュニケーションに安心と楽しさを付加するソフトウェア — FaceCommunicator<sup>®\*1)</sup> —

宮崎 敏彦      井上 清司  
天本 直弘      石川 和弘

ADSLや光ファイバの普及、さらには第三代携帯電話や無線LANの普及など、ネットワークのブロードバンド化とパーソナル化が進んでいる。このような高速大容量かつ場所と時間を選ばないというネットワークの特性を有効に活用し、かつ、その普及をさらに促進するクライアントアプリケーションのひとつとしてTV電話やビデオチャットのような、人と人のコミュニケーションを目的とした映像通信サービスが期待されている（以下ではこれらを総称してビジュアルコミュニケーションサービスと呼ぶ）。

しかし、ビジュアルコミュニケーションサービスには次のような課題があり、普及の障害のひとつになっている。

## ① プライバシー保護

唐突に掛かってくる電話の場合、例えば髪が乱れているので出たくない、あるいは部屋が散らかっているのを見せたくない等、ありのままの状態が常に写ってしまうことに対する抵抗感がある。一方で、知り合いからのTV電話による呼び出しに対し、ことさら顔を出さないで応答するのも失礼な感があり、いちいち顔を出せない理由を言うのも面倒である。

## ② 利用場面の不足

知っている者同士の場合、単に顔が見えるというだけでは面白味に欠け、顔を出す必然性が無いことも多い。

上記のような課題を補うために、沖電気では、動画像処理による顔の表情解析技術と3次元コンピュータグラフィックス（以下3DCG）による顔アニメーション技術を融合したソフトウェア“FaceCommunicator”を開発している。FaceCommunicatorは、発話者の顔画像や音声あるいは入力されたテキスト等を解析し、その内容に合った顔アニメーションを生成するソフトウェアである。FaceCommunicatorが組み込まれたビジュアルコミュニケーションサービスでは、相手の画面上に自身のそのままの顔を表示する代わりに、顔の動きを忠実に再現した3DCGのキャラクタのアニメーションを表示することによって会話することが可能となる。キャラクタの顔が発話者の顔の動きに同期することにより、こちらの表情を聞き手に伝えつつ、プライバシー保護にも配慮したビジュアルコミュニケーションサービスを実現することができる。また、アニメーションであることを活用し、その日の気分や話題に応じたキャラクタ（例えばクリスマス日にはサンタクロースのキャラクタ）を使う、あるいは実際の顔では作れない大袈裟な驚き顔を見せるなど、映像通信を積極的に利用し、その上でさまざまな遊びも提供することが可能となる（図1参照）。

本稿では、最初にFaceCommunicatorのソフトウェア

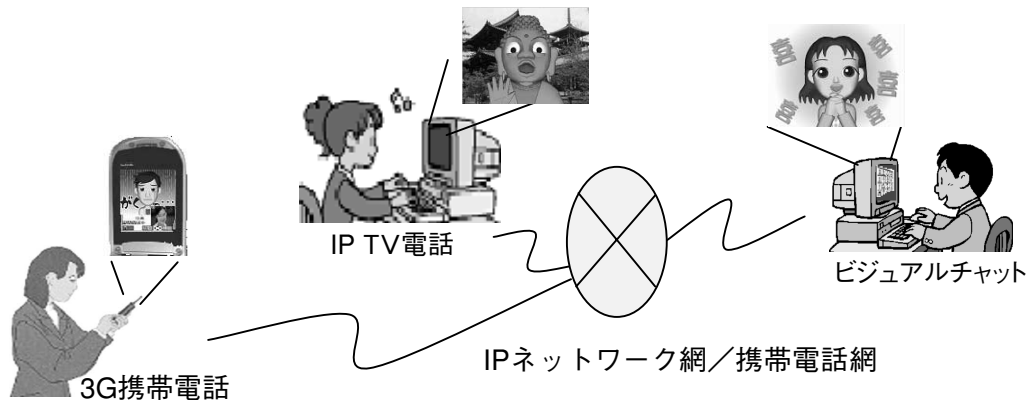


図1 FaceCommunicatorを使ったビジュアルコミュニケーション

\*1) FaceCommunicatorは、沖電気工業(株)の登録商標です。

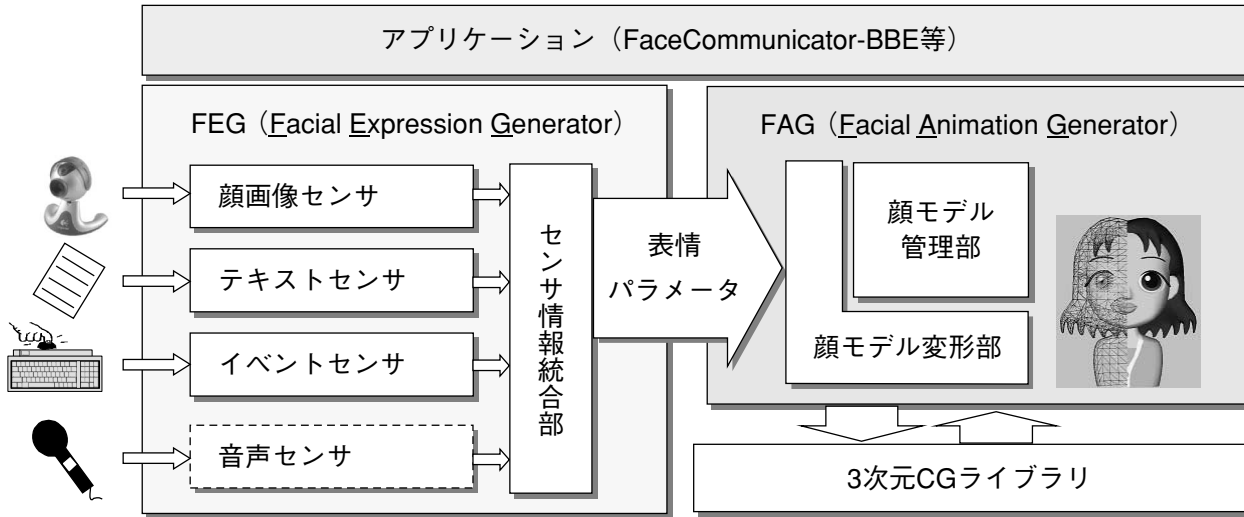


図2 FaceCommunicatorのソフトウェア構成

構成と各部の機能について概説する。次に、PCを使ったインターネット市場と携帯電話等の組込市場をターゲットとした2つの製品について、その特徴を簡単に説明する。

### FaceCommunicatorの概要

先に述べたように、FaceCommunicatorは通信相手に3次元グラフィックスを使った顔アニメーションを見せることによって、ユーザに新しいビジュアルコミュニケーション環境を提供するソフトウェアである。顔アニメーションの生成は、三次元の形状データとして表現された顔のモデル（以降単に顔モデルと呼ぶ）の目や眉あるいは口等の各部位を、必要に応じて随時変形あるいは移動することにより行われる。3次元グラフィックス技術を使った表情アニメーションは、既に映画などで盛んに使われている技術であり、ここではその手法を詳しく述べることはしないが、顔モデルの各部位の変形量を適切に与えることにより、顔モデルの上に意図通りの表情を作ることができる。この変形量のことを以下では表情パラメータと呼ぶ。

図2にFaceCommunicatorの概略的なソフトウェア構成図を示す。図において左に示したのがFEG（Facial Expression Generator）と呼ぶ表情パラメータ生成部である。また、FEGが生成した表情パラメータを入力として、顔モデルを変形し、最終的なアニメーションを生成するのが右のFAG（Facial Animation Generator）である。FEGとFAGはライブラリとして実現されており、このライブラリを使ってIPテレビ電話やチャットなど用途に応じたFaceCommunicatorを構築することができる。

FEGは、入力情報を解析する複数のセンサ部と、これ

らセンサ部から得られる情報を取捨選択あるいは統合し最終的な表情パラメータとして出力するセンサ情報統合部からなる。現在は、顔画像センサ、テキストセンサ、イベントセンサが実現されており、音声センサも開発中である。

以下に各センサの働きについて簡単に説明する。

#### (1) 顔画像センサ

顔画像センサとは、カメラに写った発話者の顔をリア

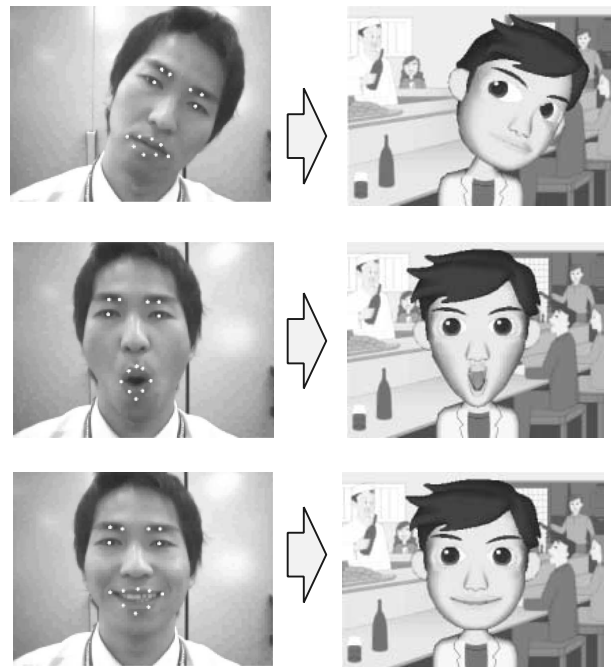


図3 顔画像センサーによるアニメーション例

ルタイムに解析し、目や眉、口など顔の特徴点の動き情報を抽出するセンサである。このセンサの働きにより、例えば発話者が右を向けばアニメーションの中のキャラクタが右を向き、発話者が眉根を寄せて怒った顔をすればキャラクタの眉根も寄って怒ったような顔をするといった、人間の顔の動きに連動した顔アニメーションを生成することができる。図3に特徴点の抽出とそれを反映したキャラクタの表示例を示す。また、顔モデルの作り方を工夫することによって、発話者の動きを忠実に再現するだけでなく、例えば犬のキャラクタを使った場合に、話者が首を左右に振ると犬のキャラクタが尻尾を振るといったように、顔の動き情報をキャラクタ上の別の動きに置き換えることも可能である。

沖電気の顔画像センサの特徴は、携帯電話など組込システム上でも動作するよう、QCIF (176×144) サイズの小さな顔画像でも顔の特徴点を抽出できる点にある。

## (2) テキストセンサ

テキストセンサとは、入力されたテキストを合成音声で再生すると同時に、再生している音声の音韻情報を元にキャラクタの口を動かすアニメーションを生成する機能である。入力するテキストの中に表情を変えるようなコマンドを文字列として入れることにより、発話中のキャラクタの表情を好みに応じて変えることもできる。例えば“(^o^)”という文字列が、『笑った顔に変える』という意味の表情コマンドであった場合に、“(^o^) こんにちは”のような文字列を与えると、キャラクタが笑った顔で「こんにちは」と発話するといった具合である。勿論、テキストの途中で何度でも表情コマンドを指定できるため、それを応用して、顔アニメーション付きのテキストチャットやメール読み上げを実現することができる。

## (3) イベントセンサ

イベントセンサはキーやマウスボタンの押下など、予め設定された特定のイベントの発生に合わせ、表情パラメータを生成するセンサである。これは図4に示すような定型アニメーションを再生するために使われるセンサであり、ユーザは会話の状況や文脈に合わせて『困る』『笑う』『謝る』等の定型アニメーションを任意の時点で再生することができる。なお、定型アニメーション自体は顔モデルが持つデータであり、クリエイターの工夫により個々の顔モデルのキャラクタ設定に合った定型アニメーションを用意することができる。

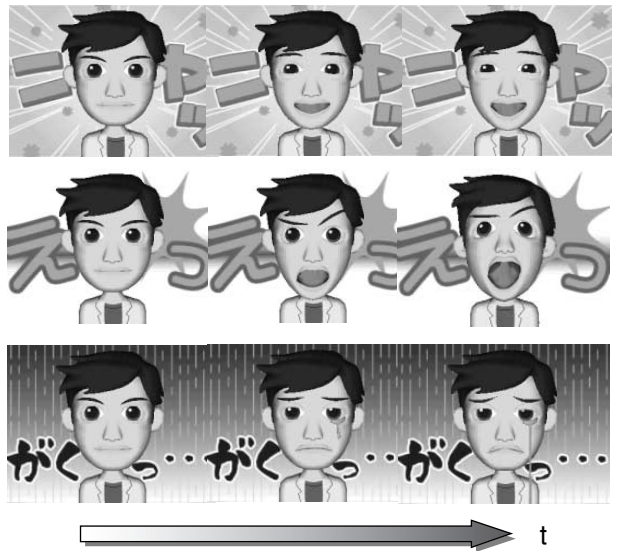


図4 定型アニメーションの例

## (4) 音声センサ

音声センサは発話者の音声を分析し、そこから得られる音韻情報を基に顔モデルの口の動きに相当する表情パラメータを生成するセンサであり現在開発中である。音声センサを使えば、カメラを持たないユーザが音声だけで相手の画面にアニメーションを表示するといったことも可能である。

## (5) センサ情報統合部

センサ情報統合部では、複数のセンサから出力される表情パラメータを統合あるいは取捨選択しFAGに渡す。センサ情報をどのように統合して使うかは、優先順位による方法とセンサから得られる確信度による方法がある。優先順位を使った例では、テキスト読み上げ時に、顔全体の向きや眉の動きは画像センサの出力を使い、口の動きのみテキストセンサからの出力を使うといった例が考えられる。確信度を使った例では、例えば逆光などカメラの画質が悪く画像センサが正しく顔の特徴量を抽出できない場合に、音声センサを使うことにより、顔モデルの口の動きを発話者の音声に同期させるといった使い方が考えられる。

## FaceCommunicator関連製品

FaceCommunicatorは、FEGとFAGの機能を使った製品の総称（製品ブランド名）である。具体的な製品としては以下のものがある。

### (1) PC用ビデオキャプチャドライバ

Windows<sup>®\*2)</sup>上のIPテレビ電話ソフトウェアやビジュアルチャットソフトウェアをターゲットとして開発しているのが、FaceCommunicator-BBE (Broad Band Edition) (以下単にBBE)である(図5)。Windows上の最近の映像通信ソフトウェアではほとんどの場合、動画画像や音声を扱う部分にマイクロソフト社が提供しているDirectShow<sup>®\*2)</sup>というライブラリが利用されている。DirectShow<sup>®</sup>ではキャプチャデバイスやエンコーダ、デコーダあるいはネットワーク通信といった論理的にまとまった部品をフィルタと呼び、必要な機能のフィルタを適宜組み合わせることによって比較的容易に目的のアプリケーションを開発できる。

BBEは、DirectShow<sup>®</sup>のキャプチャデバイスフィルタとして認識されるように作られており、BBEをインストールすると、映像通信アプリケーションのカメラ選択オプションの中に新しい選択肢としてBBEが現れる。ここでBBEを選択すると、以降、あたかもカメラが顔アニメーション機能を持つように利用することができる。このようなアーキテクチャを採用することにより次のようなメリットが生まれる。

仮想的なカメラを選択するだけであるため、既に映像通信サービスを利用しているユーザが新たにBBEを使用する場合でも、その映像通信ソフトウェアを変更することなく、BBEを使ったキャラクタのアニメーションを通信相手に送ることができる。

BBEは、カメラからの入力映像をキャラクタのアニメーション映像に置き換えているだけであり、映像や音声の通信部分は、既存の映像通信ソフトウェアを利用する。し

たがって、通信の受け手がBBEを持たない場合にも、受け手から見ると従来通りのストリーミングデータが送られてくるだけであり、受け手側で顔アニメーションを再生するための特別な仕組みは必要無い。まずは使いたい者だけが導入するだけでサービスが成り立つという性質は、新しいサービスを始める上で非常に重要である。

### (2) 組込用FaceCommunicator開発キット

FEGおよびFAGは、携帯電話やセットトップボックスといったコンシューマ向け専用装置のように、限られたリソースとCPU性能の装置においても十分な速度で動作するように設計されたソフトウェアである。これらの装置では、浮動小数点演算が高速で利用できないプラットフォームも多く、画像センサ部やアニメーション生成部はこのことを考慮して実現されている。これらFEGとFAGを組込アプリケーション開発キットとしてまとめたものをFaceCommunicator-MME (Mobile Multimedia Edition)と呼び、コンパイル用のヘッダファイルとライブラリで構成される。

顔画像の解析性能と顔アニメーションの描画性能はターゲットとする組込システムのアーキテクチャにも依存するため一概に言うことが難しいが、100MHzで動作するARM<sup>®\*3)</sup>9での動作実績がある。

## あ と が き

本稿では、今後の成長が期待されるビジュアルコミュニケーションサービスの付加価値機能として、安心と遊びを提供するソフトウェア、FaceCommunicatorについて概説した。今後FaceCommunicatorがユーザを獲得していくためには、ソフトウェアの普及は勿論のこと、利用できる顔モデルの充実が重要であり、顔モデルの作成環境の充実化を図っていく予定である。 ◆◆

### ● 筆者紹介

宮崎敏彦：Toshihiko Miyazaki. 金融ソリューションカンパニー ITインキュベーション本部 画像ソリューション開発部

井上清司：Seiji Inoue. 金融ソリューションカンパニー ITインキュベーション本部 画像ソリューション開発部

天本直弘：Naohiro Amamoto. 金融ソリューションカンパニー ITインキュベーション本部 画像ソリューション開発部 スマートセンシングチーム

石川和弘：Kazuhiro Ishikawa. 金融ソリューションカンパニー ITインキュベーション本部 画像ソリューション開発部 画像変換チーム

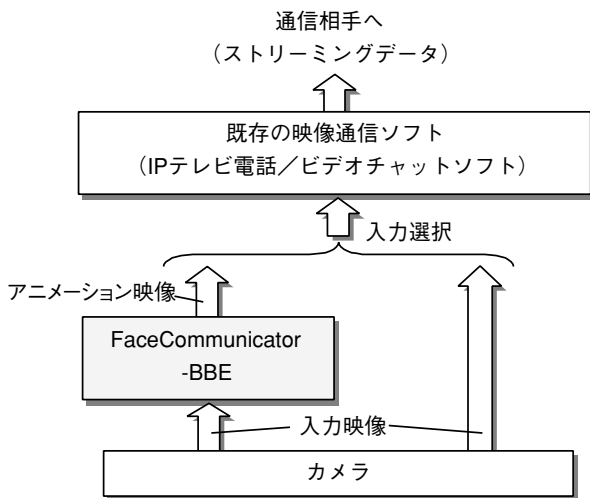


図5 FaceCommunicator-BBEと映像通信ソフト

\*2) Windows, DirectShowは米国Microsoft Corporationの米国およびその他の国における登録商標です。

\*3) ARMは、ARM社の登録商標です。