

音声に反応する表情合成システム「ニューロベビー」

4 N-9

柿本 正憲* 土佐 尚子** 森 淳一*** 真田 麻子***

*富士通研究所 **武蔵野美術大学, 総合電子専門学校 ***ビッグタウンズ

1 まえがき

顔の実時間表情合成の研究¹⁾²⁾³⁾が盛んだが、筆者らは人間の感情モデルを導入し、人間の音声に対して、CGで作った赤ちゃんの表情とCDからサンプリングした赤ちゃんの声によって応答するシステム「ニューロベビー」を試作した。ニューロベビーに対し人間がマイクを通して呼びかけると、その声の調子に応じてニューロベビーが反応する。例えば人間が楽しい声を出すとニューロベビーは笑った顔と笑い声によって反応する。怒った声を出すとニューロベビーも怒った顔と怒った声で反応する。

2 感情モデル

ニューロベビーでは、人間の感情を次のようにモデル化している。仮定として、声の調子と顔の表情に共通に対応付けることのできる感情が、多次元の距離空間である「感情座標系」を形成すると考える。この仮定に従えば、喜怒哀楽などの感情のタイプはそれぞれ感情座標系の特定の領域に配置できる。実際のシステムでは、感情座標系の二次元平面への投影を考える。以後この投影された平面座標系を単に感情座標と呼ぶことにする。また座標値(x, y)を感情変数と呼ぶ。

3 システム構成

ニューロベビーのシステム構成を図1に示す。ニューロベビーは、(1)音声解析部、(2)画像生成部、(3)音声出力部の三つのサブシステムからなる。音声解析部から画像生成部及び音声出力部に対しては、感情変数(x, y)の受渡しが行なわれており、連続的な音声入力に反応できるようになっている。

音声解析部では人間の音声波形を分析してニューラルネットワークに与え、その出力として感情変数を得る。ニューラルネットワークには予め人間の声の調子と感情変数との関係を学習させてある。

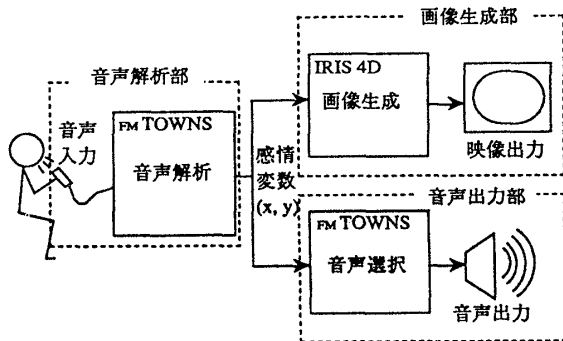


図1 ニューロベビーのシステム構成

画像生成部では感情変数を入力として赤ちゃんの表情の絵を出力とする。いくつかの感情変数に対応する顔は、予めデザインして感情座標上に配置しておく。実行時には、入力された感情変数の値に応じて、予めデザインした顔の合成を行ない表情を生成する。

音声出力部では、入力された感情座標の値に応じて、あらかじめサンプリングした音声を選択して出力する。

ハードウェアとして、音声解析部と音声出力部にはそれぞれFM TOWNSを、画像生成部にはIRIS 4Dを用いている。

4 ニューロベビーのデザイン

4-1 音声解析部のデザイン —音声と感情の関係の学習—

音声と感情との関係については文献³⁾などの研究があるが、筆者らは、ニューラルネットワークの学習機能を利用して音声から感情を抽出する手法を開発した。

人間の音声と感情変数(x, y)との関係の学習方法を図2に示す。まず、感情のタイプをいくつか設定し、感情座標に配置する。現在は感情タイプとして、喜怒哀楽の四種類を用いている。次に、それらの感情タイプを表現していると思われる典型的な声を決め、それぞれについて複数の人間の声をサンプルする。さらに、サンプルした波形から特徴量を抽出する。特徴量としては、ある時間区間における波形の最大振幅aと、ゼロ交差回数zを用いる。

ニューラルネットワークには、続けて何回か抽出した特徴量(a, z)と感情変数(x, y)との関係を学習させる。学習アルゴリズムはバックプロパゲーション法⁴⁾を使用している。

4-2 画像生成部のデザイン —顔の表情のモデリング—

画像生成部では、感情変数を入力とし、その感情に適した赤ちゃんの表情を合成して画面に絵として表示する。表情は、音声解析部で学習させた感情座標上の喜怒哀楽に対応するように三次元モデラによって作成した。これを図3に示す。顔に変化

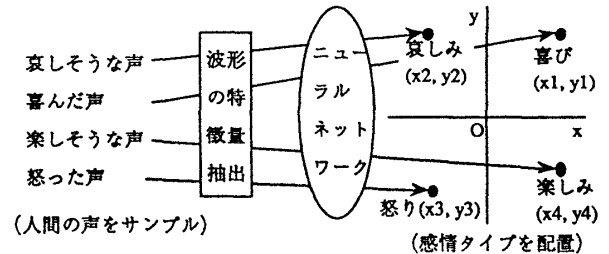


図2 音声と感情の関係の学習

Neuro-Baby — An automatic facial expression synthesizer that responds to human voice

Masanori Kakimoto¹, Naoko Tosa², Jun-ichi Mori³, Asako Sanada³

1 Fujitsu Laboratories, 2 Musashino Art University and Total House of Education for Media and Computer, 3 Big Towns

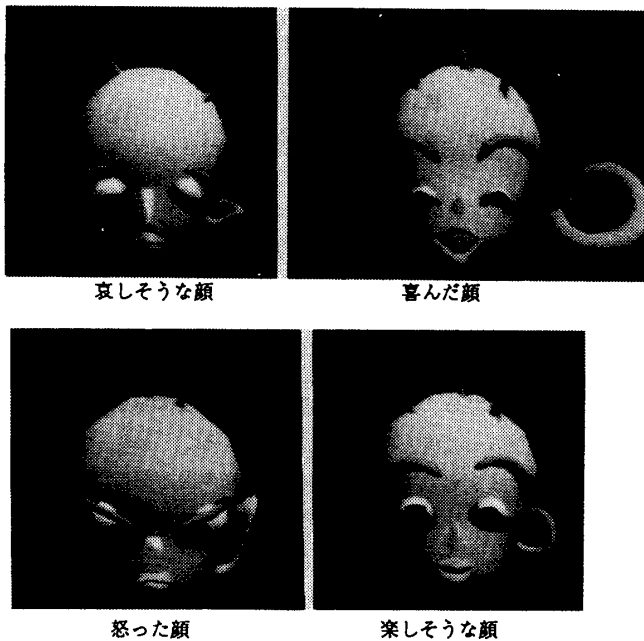


図3 デザインされた顔の表情

を持たせるために、このほかにも四種類の顔をデザインして感情座標に配置した。

表情は、眉毛・口・目などの顔のパーツの形状と配置によって決定される。パーツはそれぞれ多数のポリゴンからなる。パーツの形状モデリングには自由形状変形¹⁾を用いた。

4-3 音声出力部のデザイン

音声出力部では、使用許諾を得た市販のCD²⁾から各感情タイプに適した音声や効果音を約20種類サンプルし、感情平面上に作った領域に対応付けた。

5 ニューロベビーの実行処理

5-1 音声解析部の処理 —感情変数の生成—

音声解析部の処理は、まずマイクから取り込んだ音声波形から特徴量を計算する。これは学習の段階で行なったやり方と同じである。特徴量を受け取ったニューラルネットワークは感情変数(x, y)を出力する。感情変数は1~2秒に1回出力される。感情変数の急激な変化は現実感を損ねるので、一回の感情変数の変化量は一定の距離以内に押える。距離を調整することにより、個々のニューロベビーに、感情を押える性格とか感情の起伏の激しい性格などの個性を持たせることも可能である。

このようにして得られた感情変数は画像生成部と音声出力部に送られる。

5-2 画像生成部の処理 —表情のアニメーション生成—

ニューロベビーの実行時には、感情変数(x, y)の値に応じて、実際にデザインした各表情を重み付けして合成³⁾したような表情が出力される。具体的には、(x, y)からデザイン時に顔を配置した(x_i, y_i) (i=1, 2, ..., n)への距離をそれぞれ計算し、距離に反比例した重みを付けて各タイプの表情を合成する。実際には各感情タイプに対するパーツの頂点座標が、求められた重みに従って合成され正規化される。

一方、音声解析部からの感情変数は1~2秒に1回の頻度でしか送られて来ない。そこで、顔の表情が滑らかに変化するように見せるために、頂点情報を時系列方向に補間して顔の各パーツの形状を求め、アニメーションとして表示している。

5-3 音声出力部の処理 —効果音の選択—

音声出力部では、入力した感情変数の値に応じ、4-3で述べたように予めサンプルした2~3秒間の音声を選択して出力する。

5-4 イベント処理 —感情の時間的推移の考慮—

感情変数の時間的な推移に関していくつかの条件を設け、その条件に当てはまる場合に、予めデザインした特定の表情や音声を出すことができる。このような反応をイベントと呼んでいる。例として以下のようなものがある。

- 泣く（「哀」の付近に一定時間以上留まった場合）
- 口笛を吹く（一定時間以上音声入力がない場合）
- あくびをする（原点付近に一定時間以上留まった場合）

イベント処理は、感情座標平面に時間軸を加えて三次元の感情座標空間を扱っていると考えることができる。

6 評価

現在のところ、音声解析部での処理に安定性が低いという問題がある以外は予想以上の成果が得られた。

このシステムは、91年11月に東京・青山の小原流会館で開かれた「移動する遊体展」で一般に公開された。反応が鈍いなどの評価も中にはあったが、多くの人から好評を得た。女性、特に年配の女性はこのシステムで赤ちゃんの表情をコントロールするのが上手であった。

7 まとめ

ニューロベビーのシステムの概要について述べた。今後は、映像の質を上げるとともに、より安定した反応が得られるように学習データや学習方法の改良等を進める。また、タッチスクリーンを使った入力や、顔以外の身体による表現も計画している。将来は子供の情操教育などへの適用も考えている。

謝辞 本研究の機会を与えて下さった富士通研究所の森田修三基盤システム研究部門長代理に感謝いたします。また、議論を通じ多くの有意義な示唆・指導をして下さった大阪学院大学の太田皓一教授、音声に関してアドバイスを下さった富士通研究所の松本達郎研究員に感謝いたします。

参考文献

- [1] 森島, 原島, "テキストおよび音声に基づく顔表情自動合成のためのリアルタイムアニメーションシステムの構成," 第6回NICOGRAPH論文コンテスト論文集, pp.144-152, 1990年11月.
- [2] 大場, "多重内挿法:MIME (Multiple Inbetweening Method)によるアニメーション表現," 第6回NICOGRAPH論文コンテスト論文集, pp.153-160, 1990年11月.
- [3] 北原, 東倉, "音声の韻律情報と感情表現," 電子情報通信学会音声研究会技術報告, SP88-158, 1989年3月.
- [4] Rumelhart, Hinton, Williams, "Learning Internal Representations by Error Propagation," *Parallel Distributed Processing*, vol. I, pp.318-364, MIT Press, 1986.
- [5] 柿本, 村上, "軸を利用した自由形状変形の制御," 情報処理学会グラフィクスとCAD集中研究集会研究報告, 91-CG-52-8, 1991年8月.
- [6] Philips, Sound Ideas, "Sound Effects Library LHH7910 Series," CD1001-CD1028.