

高速度カメラによる動的な顔面表情の分析および合成

TETSUTANI, Nobuji / 四倉, 達夫 / 内田, 英子 / 山田, 寛 / 赤松, 茂 / 鉄谷, 信二 / 森島, 繁生 / YOTSUKURA, Tatsuo / UCHIDA, Hideko / YAMADA, Hiroshi / AKAMATSU, Shigeru / MORISHIMA, Shigeo

(出版者 / Publisher)

電子情報通信学会

(雑誌名 / Journal or Publication Title)

電子情報通信学会技術研究報告. HCS, ヒューマンコミュニケーション基礎 / 電子情報通信学会技術研究報告. HCS, ヒューマンコミュニケーション基礎

(巻 / Volume)

101

(号 / Number)

610

(開始ページ / Start Page)

7

(終了ページ / End Page)

12

(発行年 / Year)

2002-01-25

高速度カメラによる動的な顔面表情の分析および合成

四倉達夫¹・内田英子²・山田 寛³・赤松 茂^{4,5}・鉄谷信二⁶・森島 繁生¹

¹成蹊大学・²サンフランシスコ州立大学・³日本大学・
⁴法政大学・⁵国際電気通信基礎技術研究所・⁶ATR メディア情報科学研究所

¹ 〒 180-8633 東京都武蔵野市吉祥寺北町 3-3-1

E-mail: yotsu@ee.seikei.ac.jp

あらまし 本稿では、人間が自然な表情した場合：自発表出と典型的な表情を演じる際の顔表情：演技表出を撮影し、顔の各部位に設定した特徴点の変位量に基づき顔の動きの定量的な測定を高速度カメラを用いて分析した。また測定結果からCGによって構築した顔モデルのアニメーション生成を行った。自発表出条件、演技表出条件ともに顔の各部位の動き出しの差は微細であり高速度カメラを用いたことの有効性が示された。また情動ごとおよび表出条件ごとに顔の動き量や速さに特徴的な違いが認められたが、動きの変化そのものの様相には興味深い共通性が認められた。顔モデルのアニメーションに関しても、線形補間によるキーフレームアニメーションと比べより自然な顔表情表出が可能となった。

キーワード 顔面表情、自発表情、演技表情、顔合成、高速度カメラ

Analysis of Facial Movements and Synthesis Spontaneously Elicited and Posed Expressions of Emotion Using High-Speed Camera

Tatsuo YOTSUKURA¹, Hideko UCHIDA², Hiroshi YAMADA³,
Shigeru AKAMATSU^{4,5}, Nobuji TETSUTANI⁶, and Shigeo MORISHIMA¹

¹Seikei University, ²San Francisco State University, ³Nihon University,
⁴Hosei University, ⁵ATR International, ⁶ATR Media Information Science Laboratories

¹ 3-3-1 Kichijoji-Kitamachi Musashino-shi Tokyo, Japan, 180-8633

Abstract The present study investigated the dynamic aspects of facial movements in spontaneously elicited and posed facial expressions of emotion. We also simulated facial synthesis by using results from an analysis, the animations that we produced confirmed differences in the intensity of the facial expressiveness. We recorded participants' facial movements when they were shown a set of emotional eliciting films, and when they posed typical facial expressions. We measured facial movements frame by frame in terms of displacements of facial feature points. Such micro-temporal analysis showed that, although it was very subtle, there exists the characteristic onset asynchrony of each part's movement. Furthermore, it was found the commonality of each part's movement in temporal change although the speed and the amount of each movement varied along with expressional conditions and emotions.

Key words Facial Expressions, Elicited and Posed Expressions, Facial Synthesis, High Speed Camera

1 はじめに

近年の顔画像処理技術の発展により、仮想空間上でのアバタを介したフェイス・トゥ・フェイスコミュニケーション、エージェントとのインタラクションシステム、そして映画やアミューズメントなどのエンターテインメント分野など多岐にわたり利用され、今後の更なる需要の拡大が期待される [1]。アバタやエージェントの生成に関してはさまざまな技術があり、現在販売されている商用ソフトウェアの中にも顔モデル生成・アニメーション用アプリケーション、プラグイン等数多く存在している。

ところで、上述したような顔アニメーションソフトウェアや心理学実験に対して常に問題にされるのが、人間の実際の表情表出にかかわる顔の動きが、モルフィング技法で表現されるような線形的な動きではないのではないかという点である。現在アニメーション生成での一般的な手法としてアニメーションの開始フレームと終了フレームの顔モデルを決定し、その間の動画像を線形的な動きとして仮定した後線形補間を行い中間フレームの顔モデルを決定するものがある。しかしながらこの手法だと自然な顔表情の表出が難しく熟練したアニメーターが経験的に中間フレームの顔の各器官の変化量を決定することが多い。

また、上述したような問題ばかりでなく、演技や故意で作られたような表情が、どの程度、自然な状況の中で起こっているのか、という問題もよく指摘されるところである [2]。しかし、そのような方法を現時点では採用せざるを得ない。その理由として、これまでのところ人間の自然な情動の表出にかかわる顔の動きを定量的に記述したような研究は皆無に等しいからである。

心理学の分野において、日常生活の中で人間が見せる自然な表情を取り扱った研究はある (Fernandez-Dols & Ruiz-Belda [3])。また、なんらかの情動喚起刺激を被験者に提示し、それによって被験者が自発した表情を調べるという研究も行われている。Ekman et al.[4] は、被験者に情動を喚起させるような映画刺激を提示し、その際に被験者が顔に表した表情を FACS (Facial Action Coding System) [5] を用いて記述している。このように自然な自発表情の分析に関する研究はこれまでも行われて来てはいるが、自然な表情にかかわる顔の動きをコンピュータ・グラフィックスで再現するために必要とされるような詳細なデータを提供してくれる研究は見受けられない。

そこで、本研究では、人間が自然に表情を自発させる時の顔の動きおよび典型的とされるような表情を演じる時の顔の動きを微細に記録し、その結果から顔モデルのアニメーションを行うことを目的とする。すなわち、いくつかの基本的な情動を喚起させることができるとされる Gross & Levenson [6] が作成した映像・映画刺激を被験者に提示し、それに対する被験者が自発的な顔の動きを 250 フレーム/秒の高速度カメラで撮影した。また、そ

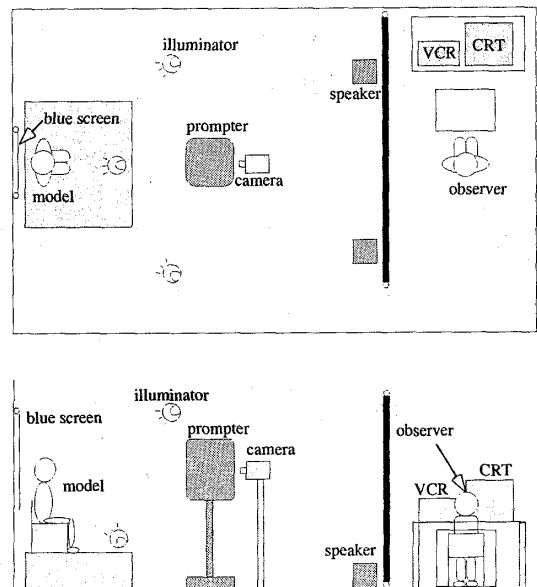


図 1: 実験装置類の配置図

れとともに、Ekman & Friesen [5] が定義しているような典型的で普遍的な表情を演じる際の顔の動きも同様に撮影した。さらに、そのようにして得られた動画像から、顔の表情を合成するために必要とされる顔の特徴点の変位量を測定し、自発表情および演技表情における顔の動的变化の特徴について分析を行った。その結果をもとに表情変形可能な顔モデルに各フレームでの顔部位の変化量を割り当てアニメーションの生成を行い、線形補間によるアニメーション生成および非線形補間でのアニメーション生成の違いを検証した。

2 方法

2.1 被験者

被験者は ATR の被験者プールの中から、モデル経験のある大学生および社会人 84 名を抽出し、面接を通じて 24 名 (男 12 名、女 12 名) を選定した。選定の基準は 20~35 歳の日本人であること、表情が豊かであること、口髭やあご髭がないこと、視力が裸眼もしくはコンタクトレンズによる矯正で 1.0 以上であることとした。

2.2 実験装置

本実験では、被験者の自発表情と演技表情の動的な変化を高速度カメラ (Nac : SV-500c3) で撮影し S-VHS ビデオに録画した。録画速度は 250 [fps]、シャッター速度は 1/250 [sec] であった。なお、被験者と高速度カメラの間には、プロンプター (キャノン : CWP-10H・II) をおいた。これは、被験者の顔を正面から撮影しながら、被験者に自分自身の顔の動きや映像刺激を提示するためである。以上の実験装置類を配置した実験室の様子を図 1 に示す。

2.3 情動喚起刺激

自発表情を撮影するセッションで Gross & Levenson[6]により標準化された情動喚起刺激セットを使用した。これは、喜び、驚き、怒り、悲しみ、嫌悪、恐れ的情動が喚起できたと実験的に評価された映画のワンシーンのクリップや彼らが独自に撮影した映像クリップで構成されている。各クリップの合間には、ニュートラル(無感情)な映像をはさみ、各情動喚起シーンがランダムな順序で提示されるように刺激ビデオテープを編集した。

2.4 手続き

実験は ATR 内の実験室において被験者ごとに個別に実施した。実験は 3 つのセッションに分けて実施した。はじめに 1) トレーニングセッション、続いて、2) 自発表情撮影セッション、最後に 3) 演技表情撮影セッションの順番で行った。各セッションの詳細を以下に示す。

2.4.1 トレーニング・セッション

被験者に実験環境に慣れてもらい、また自然な表情が出易いように、ウォーム・アップに実験者の指示により FACS (Facial Action Coding System) [5] に基づいた表情の筋肉を動かすフェイスナル・エクササイズを行った。なお、ここでは表情を作る(顔の筋肉を総合的に動かす)のではなく、表情の筋肉を単独に動かす練習を 30 分~1 時間程行った。

2.4.2 自発表情撮影セッション

本セッション(自発表出条件)では、Gross & Levenson [6]により標準化された情動喚起映像を提示し、被験者に自然な表情を自発させた。なお、このセッションでは、被験者が実験者を意識して自然な表情を抑制するというようなことができるかぎり起こらないよう配慮した。すなわち、実験者は、被験者を一人にして、被験者からは見えない実験室の裏に移動し、その場所で、情動喚起映像の提示操作やカメラの微調整などを行った。自発表出セッションの所要時間は、約 1 時間であった。

2.4.3 演技表情撮影セッション

本セッションでは、FACS をベースとした顔面動作教示に従って、被験者に 6 つの基本表情を演じさせた。ここでは、被験者に「喜んだ顔をしてください。」といった感情のキーワードを全く使わずに、表情の筋肉(アクション・ユニット(以下 AU)を顔の各部(AU ごと)に訓練し、最終的にアクション・ユニットを組み合わせて 1 つの表情がスムーズに表出できるよう訓練した。そしてリラックスした顔(ニュートラル)から基本感情が演じられるように訓練をした。なお、本セッションでは、自発表情撮影のセッションとは違い、実験者は被験者と対話しながら表情表出の訓練を行った。

リラックスした顔(ニュートラル)から基本 6 表情(喜び、驚き、悲しみ、恐れ、嫌悪、怒り)がスムーズに演じられるようになるまで表情ごとに訓練した。訓練に所要した時間は個人差はあったが、各表情につき 30 分~1 時間であった。表情表出の訓練に要した時間に個人差はあったが、所要時間は平均して 4~5 時間を要した。

3 結果

本章では先に述べた実験時に撮影した顔画像の分析方法および分析結果を述べる。

3.1 分析方法

今回の実験で、高速度カメラによって撮影し VTR に録画した 24 名の被験者の自発表出条件および演技表出条件における顔の動きを次のように処理し分析した。

3.1.1 分析用表情変化シーンの抽出

分析対象の表情変化シーンの決定条件として、自発表情撮影セッションで撮影された映像の中から、各映画刺激ごとに、各被験者が明らかにその映画刺激に対する反応として示したと判断される自発的な情動の表出シーンを選定した。演技のシーンについては、各情動ごとにそれぞれの表情が最もうまく演じられていると判断された試行で、演技開始時の無表情の状態のフレームから演技表出が最大と判断されるフレームまでとした。この分析対象シーンの選定の結果、悲しみ、恐れ、怒りの情動を誘発するとされる映画刺激を、ほとんどの被験者が表情を変化させずに観察していることが分かったため分析対象から除外した。

3.1.2 分析手順

次に、上記のようにして選出された各表情シーンにおける顔の特徴点の変位を計測した。特徴点は、左右眉輪郭各部 4 点、左右目輪郭各部 4 点、鼻輪郭部 5 点、唇輪郭部 6 点、顎部 1 点の合計 28 点とした。他に額の髪の生えぎわの 1 点(29 番)と鼻の頂点の 1 点(30 番)を特徴点位置の補正用の点として設定した。

3.2 分析結果

以上のような分析結果から 1) 各情動での顔部位の変化開始順、2) 顔部位の変化に要する時間、そして 3) 特徴点の変化量と経過時間との関係、計 3 つの結果を述べる。

3.2.1 顔部位の変化開始順

表出条件の違いによって、各情動の表出における顔の部位の動き出す順番が異なるかどうか、またそれらの時間(フレーム)差がどの程度のものかを分析した。この分析にあたっては、それぞれの情動における表出条件ごとに、一つの部位(例えば眉)に関わる特徴点の動き出しと他の部位(例えば目)のそれとの時間差の平均を求めた。これによって、表 1 に示すような結果が得られた。この

表 1: 各情動の表出条件別顔部位変化開始順位と平均時間差 (単位:ms)

情動	条件	1 番	時間差	2 番	時間差	3 番
喜び	自発	目	23.3	眉	26.1	口
	演技	目	9.2	眉	18.5	口
驚き	自発	目	5.6	口	6.0	眉
	演技	目	12.8	口	15.9	眉
嫌悪	自発	眉	0.2	口	6.4	目
	演技	目	11.6	口	4.2	眉

表 2: 各情動表出における顔の動き開始から完了までの経過時間 (単位: ms)

情動	自発	演技
喜び	724	556
驚き	228	412
嫌悪	836	388

表から、嫌悪の自発条件を除く他の条件では、全体的な傾向として、顔の中ではまず目の部分から動き始めることが分かる。また、喜びと驚きの場合には、自発条件、演技条件ともに、それぞれが、目、眉、口の順、目、口、眉の順で動いていることが分かった。特に注目すべき箇所として各々の時間差が 33[msec] より小さく、通常使用しているカメラ (30[fps]) では捕らえることができない点である。このことから自発・演技を見分ける際、通常のカメラだとこれらの時間差というパラメータが欠損する恐れがあると考えられる。

3.2.2 顔部位の変化に要する時間

顔のいずれかの部分が動き始めてから表出が完了するまでの経過時間を各情動ごとで表出条件別に平均した値を、表 2 に示す。この表から嫌悪や喜びの場合には、自発条件での表出の方が時間がかかる一方、驚きの場合は、逆に、自発条件の方が、短時間に表出が完了することが分かった。

3.2.3 特徴点の変位量と経過時間との関係

時間を関数とした各部の特徴点の変位量についての分析を行った。この分析にあたっては、まず、選出された各シーン毎に、各フレームにおける特徴点の部位別の平均変位量を求め、さらに、それらを情動ごとに表出条件別にして平均した。この結果の 1 例 (表情: 喜び) をグラフにして図 2, 図 3 に示す。なお、眉と目の動きについては左右に傾向の違いは認められなかった。このグラフから、全体的に顔の動きは、モルフィングの技法で表現されるような 1 次関数的な変化ではないことが分かる。どの情動のどの部分においても、はじめはその後と比較して変位量は少なく、どちらかと言うとゆっくりと立上り、その後急激な動きに変わることが示されている。また表出動作の完了近くになると変位量が減少し、完了に

向けて収束していくという様子が見られる。この現状は他の表情: 嫌悪、驚きの場合でも同様である。

情動: 喜びでの自発表出条件と演技表情表出条件にける動きの違いを見ると、自発表出条件と比べて演技表出条件の方が、全体的に動きが速いが、特に口が大きく速く動いていることが特徴的である。また、演技表出条件の場合には、目と眉の特徴点の変位量にほとんど違いがない。つまり、目と眉の動きの大きさに変わりがない。これに対して、自発表出条件の場合には、眉の動きにの大きさが、目の動きの大きさを上回っていることが特徴的と言える。

4 表情合成

前章までのおり表情表出時における顔の各器官の動きは線形補間で変化しているのはなく非線形で変化が生じていることが確認できた。そこで本章では分析結果をもとに顔表情アニメーションを実際に制作し、CG によって再現された顔表情においても観察者が自然に表出ができていくかどうか検証した。

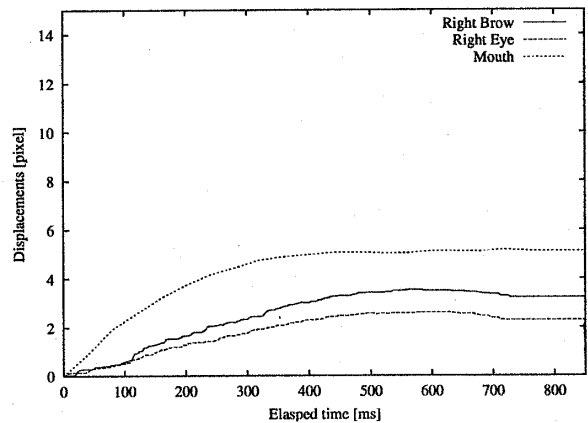


図 2: 喜びの自発表出

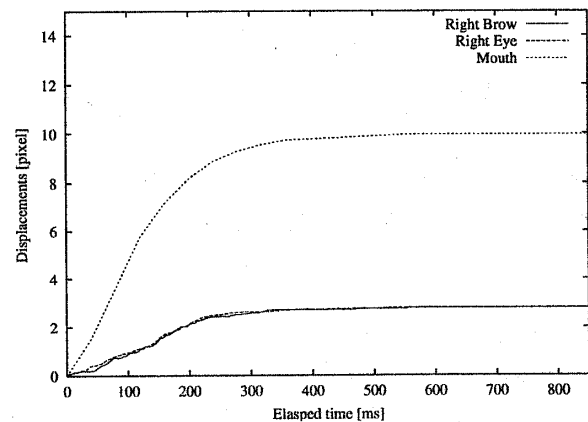


図 3: 喜びの演技表出

4.1 顔モデル合成

表情の表出が可能なリアルな顔モデルを生成するため、カメラから取得した対象人物の正面の頭部画像に三角形ポリゴンで構成させる頭部の標準ワイヤフレームモデルをマニュアル整合し顔モデルを作成する。そしてこのモデルにテクスチャマッピングを施すことによって頭部合成画像を構築する。この整合処理はマウス操作によって容易に実行することができるように、等研究室で製作した GUI ツールを用いる [1]。このツールを用いることで熟練者であれば約 1 分程度でフィッティング可能である。

4.2 表情編集ツール

上記の顔モデルに対し表情変化を行うため、FACS に基づいたルール付けを行う。基本動作である各 AU はいくつかの顔面上の特徴点の 3 次元移動ベクトルとして定義されている。表情変化は 3 次元モデルの特頂点を AU の強度によって移動させ特徴以外の格子点は特頂点の移動に基づく補間によって制御される。これらの制御はあらかじめ表情編集ツールと呼ばれるアプリケーションで各 AU の強度をスライダーバーによって制御することによって簡単に顔モデルの表情のカスタマイズが可能である。

4.3 表情アニメーション

顔表情アニメーションを製作するにあたりまず先述の表情編集ツールを用いて顔アニメーション用のシークエンス画像を生成する。アニメーションの描画フレームレートはシステムの関係上 30[fps] とする。表情表出時に使用する AU の強度と分析で求めた顔部位の変化量との関係は各表情ごとに変化量が最大な変位量を示す部位に対応する AU の強度を 100% と設定する。各フレームにおける顔部位の変化量から対応した AU の強度を表情編集ツールによって入力し、合成された顔画像を逐次生成してゆく。これら顔画像はイメージファイルとして保存し、市販の動画編集ソフトにてシークエンス画像を動画画像へ変換させアニメーションを生成する。

4.4 評価

上記手法によって生成した顔モデルの表情：喜び、驚きでのシークエンス画像をそれぞれ図 4、図 5 に示す。上図 (a) が一般的に従来から用いられている AU の強度を線形的に変化させて生成したもので、下図 (b) が分析によって求めた自発的な表出での表情変化である。現在のところ主観評価実験など定量的な分析を行っていないが、予備実験として数人の被験者に対して今回制作した自発表出条件と演技表情表出条件、線形補間、計 3 種類のアニメーションをランダムにどれが一番自然な表情表出をしているか評価したところ、自発的表出条件の分析結果をもとにして生成したアニメーションしたものが演技表情表出条件と比べ、わずかに一番自然な表情表出であると意見を頂いた。また線形補間で制作したアニメー

ションは多くの被験者が自然な表出ではなくぎこちないアニメーションであるとの評価が集まった。

5 まとめ

本研究では、演技表情と自発的表出時の顔の動きを高速度カメラで撮影し、顔の特徴点の変位の測定に基づいて顔の動きの定量的な特性を分析した。この研究でもっとも重要なキーワードである「高速度カメラ」を本研究に取り入れることで、詳細に顔の部位の動きを捕らえることができ、有効性を確認することができた。また分析結果でも述べたが顔部位の変化開始時間の順位と平均時間差が通常のカメラフレーム速度の 33ms 以下で部位間の時間差が生じている。これらは通常のカメラで捕らえることが不可能である。そのため問題になってくるのは合成顔画像を生成した場合での描画フレームレートである。本研究からもわかるとおり今回ハードウェアの関係上 30[fps] 以上のフレームレートでの描画が無理である為、制限された時間解像度で生成を行った。今後の展開として 30[fps] 以上描画可能な表示用機器を用いて合成表情アニメーションを行い、現在 TV 等で用いられているフレームレートと時間解像度を上げたものとを比べ印象の変化があるかどうか検証する必要がある。もし時間解像度を上げたものがより自然な印象を与えると判断できれば、将来高フレームレートの再生機器が登場する可能性も考えられる。

参考文献

- [1] Morishima, S. Modeling of Facial Expression and Emotion for Human Communication System Displays 17, pp.15-25, Elsevier, 1996.
- [2] Russell, J. A., & Lemay, G. 表情表出の次元的-文脈的観点. 心理学評論, 43 (2), 161-176. 2000
- [3] Fernandez-Dols, J. M., & Ruiz-Belda, M. Spontaneous facial behavior during intense emotional episodes: Artistic truth and optical truth. In J. A. Russell & J. M. FernandezDols (Eds.), The psychology of facial expression (pp. 255-274). New York, NY: Cambridge University Press. 1997
- [4] Ekman, P., Friesen, W. V., & Ancoli, S. Facial signs of emotional experience. Journal of Personality and Social Psychology, 39, 1125-1134. 1980
- [5] Ekman, P., & Friesen, W. V. Facial Action Coding System (FACS): A technique for the measurement of facial action. Palo Alto, CA: Consulting Psychologists Press. 1978
- [6] Gross, J., & Levenson, R. Emotion elicitation using films. Cognition and Emotion, 9, 89-108. 1995

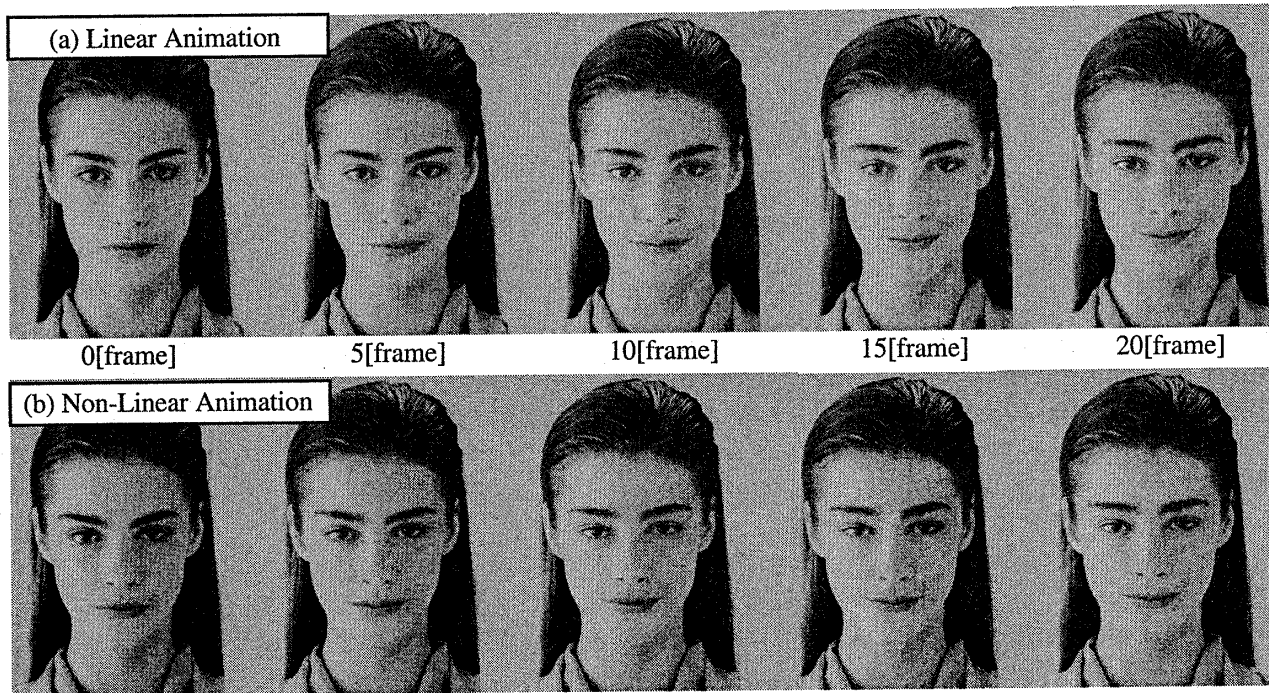


図 4: 表情 : 喜びの合成顔画像シーケンス

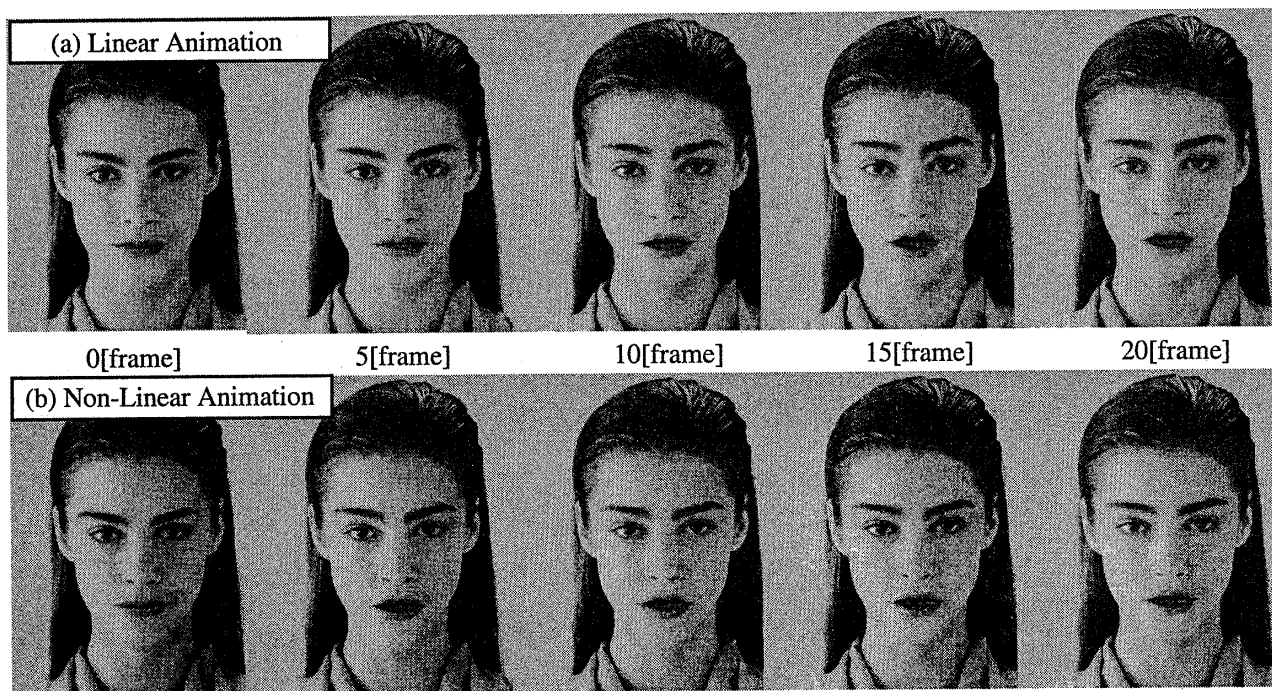


図 5: 表情 : 驚きの合成顔画像シーケンス