発話における自然な口の動作合成の一検討

4H－11

岩高 剛　　橋本 良太　　小川 均
立命館大学理工学部情報学科

1　はじめに

コンピュータ・グラフィックス（ＣＧ）による顔の動画の通信を想定する場合，画面に出力された顔の表情や口の動きなどが自然に表現できれば，あたかも実際に人間と会話をしているかのような臨場感が期待できる．そのためには，表情や口の動きの自然な画像合成が必要である

本論では，口の動きに着目し，発話における自然な口の動きの動画処理について考察する．口は発話時に，様々な形や大きさの変化が見られる．そこで，移動方向，強さ (速度)，移動距離の相互関係を調べ，自然に動作を行なう一般化した軌跡モデルを構築する．この軌跡モデルを使用して，個人特有の情報を使用せずに自然な動作を表現させることを目的とする．

対象としたのは，日本語の五母音であり，閉口状態から各々の母音を発声するまでの変化量，またある母音から他の母音へ（例えば「あ」から「い」へ）の変化量を測定し，特徴量を検出する．また，軌跡モデルを用いてＣＧの再現を行う．

2　口周囲筋による口の動作の撮影

発話に対する口の動きを調べるために,3 人の被験者 (A,B,C) に 15 個のマーカ (ポイント $P[k](k = 1 \ldots 15)$) を張り付け，30fps(frames/sec) のビデオカメラで口の動きを撮影した．解剖学分野の筋学 (Myology)[1] 見地から，口輪筋 (Orbicularis oris)，オトガイ筋 (Mentalis)，上唇挙筋 (Levator labii superioris)，口角挙筋 (Levator anguli oris)，笑筋 (Risorius)，下唇下制筋 (Depressor labii inferioris)，口角下制筋 (Depressor anguli oris) の動きが測定できるポイント (図 1) を用いた．

3　ポイントの軌跡の計測

3 人の被験者に，通常時の話し方で閉口状態から各々の母音，母音から他の母音の計 25 の動作を発話してもらい，撮影を行なった．

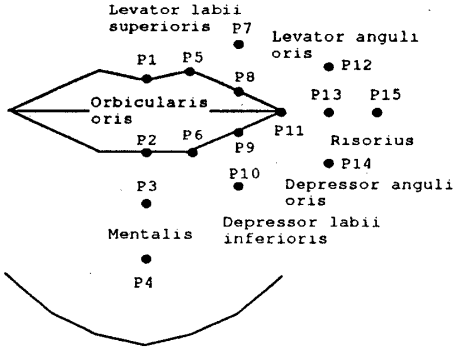

図 1: ポイントの位置

各々の動作は震えなど残るが,6 フレーム (0.2 秒) 以内に主要な動作は終えていたので 6 フレーム間の軌跡を計測した．

図 3 に，被験者 A の閉口状態から母音「あ」の各フレームにおける各々のポイントの座標値を記す．図中の矢印は動く方向を，単位は実際の移動量を示す．

4　軌跡モデルの構築

各々のポイント $P[k]$ は，図 2 のようにフレームごとに方向，及び強さ (速度) を変えながら移動する．

そこでフレーム間の方向と強さの変化を検出し，各動作の軌跡モデルを構築するため次節のように行なった．

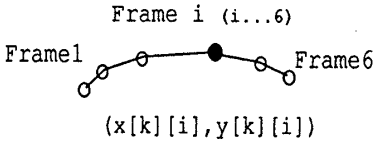

図 2: ポイント $P[k][i]$ の軌跡の例

Natural Lip Motion Synthesis in Speech
Tsuyoshi Iwataka, Ryota Hashimoto, Hitoshi Ogawa
Department of Computer Science, Faculty of Science and Engineering, Ritsumeikan University

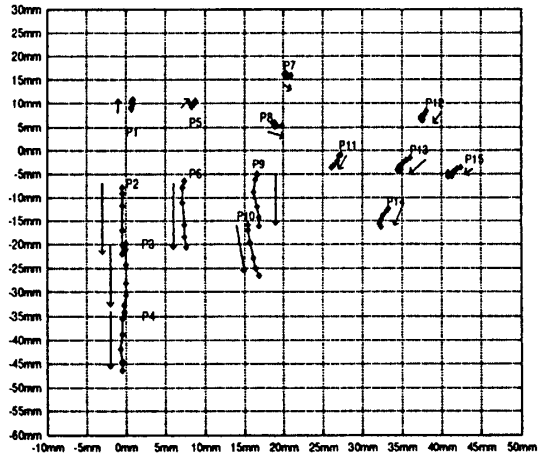

図 3: 被験者 A の閉口状態から「あ」の軌跡

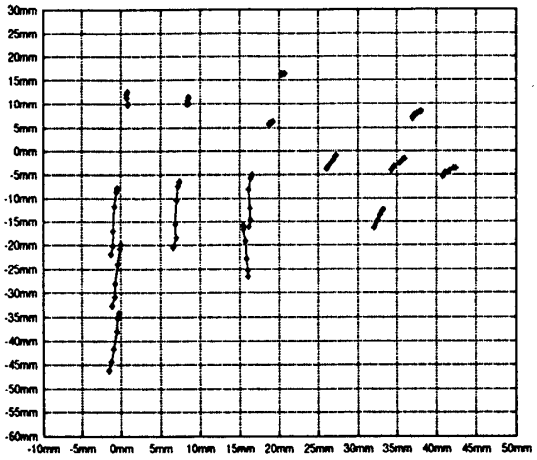

図 4: 軌跡モデルの閉口状態から「あ」の軌跡

4.1 特徴量の検出

4.1.1 移動方向の検出

各ポイントのフレーム間での方向を$\vec{e}[k][i](k = 1\ldots15, i = 1\ldots5)$で表わし, その大きさを$|e[k][i]|$とする.

$$|e[k][i]| =$$

$$\sqrt{(x[k][i+1]-x[k][i])^2+(y[k][i+1]-y[k][i])^2}$$

$$\vec{e}[k][i] =$$

$$\left(\frac{x[k][i+1]-x[k][i]}{|e[k][i]|}, \frac{y[k][i+1]-y[k][i]}{|e[k][i]|}\right)$$

3 章で得た各々の軌跡をみると, 被験者 3 人が同一の言葉を発話するとき, ポイント $P[k]$ の移動方向に類似性が見られた. よって,3 人の方向ベクトルの平均値$\vec{V}[k][i]$を求め, 軌跡モデルの方向ベクトルとする.

$$\vec{V}[k][i] = \frac{1}{3}(\vec{e}[k][i]_{(A)} + \vec{e}[k][i]_{(B)} + \vec{e}[k][i]_{(C)}) \quad (1)$$

4.1.2 フレーム間の強さの検出

ポイント $P[k]$ の総移動距離を $S[k](k = 1\ldots15)$ とすれば

$$S[k] = \sum_{i=1}^{5} |e[k][i]|$$

となる.$S[k]$ は開口の大きさを表わし, 発話時における強弱 (感情など) や個人差に大きく影響する. それらの依存関係を無効にするため,$S[k]$ に単位長の正規化を施し, フレームごとの強さ $l[k][i]$ を求めた.

$$l[k][i] = \frac{|e[k][i]|}{S[k]}$$

また,3 人の平均値 $L[k][i]$ は

$$L[k][i] = \frac{1}{3}(l[k][i]_{(A)} + l[k][i]_{(B)} + l[k][i]_{(C)}) \quad (2)$$

となり, 軌跡モデルのフレームごとの強さとした.

4.2 軌跡モデル

本実験では,25 種類の発話動作におけるそれぞの $\vec{V}[k][i], L[k][i]$ を検出した. 式 (1) は, 各フレームにおいてのポイントの動作方向の情報をもち, 発話動作を決定する. 式 (2) は, 各フレームにおいてのポイントの移動距離率の情報をもち,$S[k]$ によって, 開口時の大きさ, いわゆる強弱が決まる.

従って, 事前に軌跡モデルに$\vec{V}[k][i], L[k][i], S[k]$ の情報をもたせておけば, 個人特有の情報に影響されずに自然な発話動作が可能となる. ただし, 口の認識技術が必要とされる始点はあたえなければならない.

図 4 に軌跡モデルで計算した, 閉口状態から「あ」の軌跡を示す. ただし $S[k]$ は被験者 A の値を使用した. フレームごとの方向. 強さは. ほぼ同じ値を示した.

5 おわりに

本研究では,$\vec{V}[k][i], L[k][i], S[k]$ の情報だけで各発話を表現する軌跡モデルを構築することを目的とした. 軌跡モデルの方向ベクトル$\vec{V}[k][i]$ と強さ $L[k][i]$ を求めたが, 総移動距離 $S[k]$ の検出は現在, 検出中である. 従って今回の実験では, 被験者 3 人における $S[k]$ をそのまま使用した.

今後は,$S[k]$ の相互関係を調べ一般的な軌跡モデルを完成させる予定である. これによって, 個人特有の情報が全くなく強弱が自由にできる発話の動作合成システムを構築したい.

参考文献

[1] 小川 鼎三, 森 於菟, 森 富, 大内 弘 : “分担解剖学 1”, 金原出版株式会社, (1995) .