

# “ 振り ”の練習を支援するインタラクティブシステム

高橋 雅人\* 林 貴宏\*\* 尾内 理紀夫\*

\* 電気通信大学大学院 電気通信学研究科 情報工学専攻

\*\* 電気通信大学 電気通信学部 情報工学科

踊り、スポーツにおけるフォーム、手話など特定の型に沿った身振り練習を支援するインタラクティブシステムを開発した。本システムは、動作と音声を組み合わせた振りを対象とし、最終的にユーザが手本を見なくても振りができるようになることを支援し、さらにユーザに練習の退屈感を感じさせないことを目的としたものである。本システムは、ユーザが行った振りに対して DP マッチングにより採点を行い、さらにリプレイ動画で 1 秒毎にユーザの動作のどこが悪かったか、どうすればよいかを指摘する。また、練習中に映像効果や効果音を用いて退屈感を解消する。評価実験により、本システムの練習効果を確認した。

## Interactive System for Practicing “ Dance Motion with Voice ”

Masato TAKAHASHI\*, Takahiro HAYASHI\*\*, and Rikio ONAI\*

\*Graduate School of Computer Science, University of Electro-Communications

\*\*Department of Computer Science, University of Electro-Communications

This paper describes an interactive support system for practicing “ dance motion with voice ”. We propose, implement and evaluate the system. Using the system, a user can practice dance motion with voice efficiently. Our goal is to help the user to dance without watching the model action. First a user practices a dance. Second the system compares user action to model action using the DP matching and gives the user a score. Third the system replays user’s dance movie and gives the user directions to correct motion errors. The system gives video / sound effects corresponding to the score. These effects encourage the user to his practice. We have experimented and confirmed the effectiveness of the system.

### 1 はじめに

近年、生涯学習への関心が高まっており、趣味、健康、ボランティアなどに役立つ技能を学びたいと考えている人が多い [1]。これらの技能の中には踊り、スポーツ、手話など、身振りや手振りといった特定の型（本研究では動作と音声を組み合わせた特定の型を“ 振り ”と呼ぶ）を覚えることで習得できるものも多い。そこで、ユーザの振りを採点し、振りの上達に必要な情報を的確に示すようなシステムがあれば、効果的な練習ができ、個人でも振りを短期間に習得でき有用と考えられる。これまで、振りの練習を支援するシステムとして文献 [2][3] で提案されたものがある。これらのシステムの特徴は、画面上にユーザの CG と手本となる動作（モデル動作）を行う人の CG が同時に映し出され、モデル動作に合わせてユーザが体を動かすことである。なお、モデル動作を行う人の CG は文献 [2][3] ではゴーストと呼ばれ、本論文でも同様にゴーストと呼ぶことにする。振りの習得という目標からすると、最終的にはゴース

トがない状態でその動作ができなければ意味がない。しかし、これら [2][3] では、ゴーストを頼りにするあまり、ゴーストを消したときにユーザがどの程度うまく動作できるのか疑問が残る。また、歌や掛け声などを含む動作の練習を考えた場合、動作に加えて音声も採点する必要があるが、従来の振り練習支援システムで動作と音声を組み合わせて採点するものはない。また、練習の繰り返しによって生じる退屈感も問題である [4]。

以上のことから、本研究では動作と音声を組み合わせた振りを対象とし、ユーザがゴーストがない状態で、すなわち何も見ないで動作できるようになることを支援し、さらに練習の退屈感をユーザに感じさせないシステムの実現を目標とする。よって、提案システムではユーザが振りを行っている最中はゴーストを表示せずに、ユーザのみの表示にとどめる。ユーザとゴーストの表示については、CG を用いる方法ではなく、両者の実際の姿がモニタに映る“ 鏡像 ”[5]を導入し、姿勢などの CG では伝わりにくい部分をより具体的に伝え

られるようにする．ユーザが振りを一通り終わると、今行った振りを採点し、振りの良否をユーザに伝える．その後、今行った振りをリプレイ動画によって見られるようにする．このとき、半透明のゴーストを重ねることで動作が合っていたかどうかを具体的に確認できるようにする．さらに、リプレイ動画の再生中に動作の誤りに対して文字情報を使って動作を補正する指示を行う．練習における退屈感に対しては、エフェクトを用いてユーザの向上心を刺激する方法 [6] を導入し、振りの採点結果に応じて映像エフェクトや効果音を出力することでユーザに振りの上達に対する達成感を与え、退屈感を解消する．

本稿では、2章で提案システム、3章で評価実験について述べる．実験では、従来の練習方法と本論文で提案する練習方法について上達度の違いを測定した結果について述べ、4章でまとめる．

## 2 提案するシステム

本章では、本論文で提案する振り練習支援システムの構成と各部について述べる．

### 2.1 システム構成

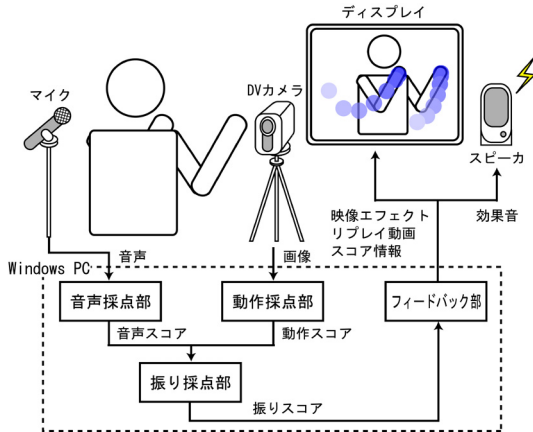


図 1: システム構成

本システムは、家庭用Digital Video(DV)カメラ1台、マイク1本、スピーカ1組(2個)、ディスプレイ1台、計算機(Pentium4-1.7GHz, Windows2000 OS)1台で構成される(図1)．カメラ映像の解像度は  $360 \times 240$  pixel, 入力時のフレームレートは 29.97fps, カメラからユーザまでの距離は約 2m である．本システムにおける処理の流れを以下に述べる(図1参照)．

- ユーザが DV カメラの前に立ち、振りを行う．
- ユーザが振りを一通り終わると、システムは今行った振りを動作と音声に分けて採点する．

動作の採点は動作採点部が行い、音声の採点は音声採点部が行い、動作と音声のスコアを計算する．これらのスコアを振り採点部が総合的に判断し、振りのスコアを 100 点満点で算出する．

- 今行った振りの良否をユーザに提示するため、フィードバック部が振りのスコアを画面に表示し、スコアに応じた効果音を鳴らす．
- 今行った振りがいつどのように誤っていたように補正すればよいかをユーザに提示するため、リプレイ動画を用いる．提示に用いるものは、ユーザ動作に対して「もっと速く」「もっとゆっくり」「もっと大きく」「もっと小さく」「素晴らしい」「もっと手本を見て」の6種類の文字情報と着色したゴーストである．
- その後続けて練習するとき、フィードバック部が練習の最中に前回の振りのスコアに応じて映像エフェクトをユーザが映し出される出力映像に出力する．

各部について以下の各節で詳しく述べる．

### 2.2 動作採点部

動作採点部は、DV カメラよりユーザ動作を取得し、ユーザ動作とモデル動作の類似度を採点し、動作のスコアとして出力する．

#### 2.2.1 動作の定量化

ユーザおよびモデル動作を定量化したものととして、本システムでは体のある箇所(マーカによって示される)の位置および移動方向の時系列  $U, V$  を用いる．具体的には、時刻  $i$  でのマーカ位置  $p_i = (x_i, y_i)$  を用いて、位置の時系列  $U$  および移動方向の時系列  $V$  を次式で定義する．

$$U \equiv [p_0, p_1, p_2, \dots, p_i, \dots, p_{N-1}]$$

$$V \equiv [v_0, v_1, v_2, \dots, v_i, \dots, v_{N-1}]$$

$$v_i = (v_{xi}, v_{yi})$$

$$v_{xi} = \begin{cases} 1: & x_{i+1} - x_i > 3 \\ 0: & |x_{i+1} - x_i| \leq 3 \\ -1: & x_{i+1} - x_i < -3 \end{cases}$$

$$v_{yi} = \begin{cases} 1: & y_{i+1} - y_i > 3 \\ 0: & |y_{i+1} - y_i| \leq 3 \\ -1: & y_{i+1} - y_i < -3 \end{cases}$$

ここで、 $N$  は動作開始から終了までの DV カメラにおけるフレーム数である．時刻  $i$  でのマー

力の移動方向を表す  $v_i$  はマーカ位置  $p_i$  から見て  $p_{i+1}$  が相対的にどの位置にあるかで決まる．図 2 を用いて直観的な説明をする．例えば，図 2 は  $p_i$  を中心として  $xy$  平面を 9 つの領域 A~I に分けたものである．各領域に書かれているベクトルは  $p_{i+1}$  が各領域にある場合の  $v_i$  を表す． $p_i = (5, 10), p_{i+1} = (9, 16)$  ならば  $p_{i+1} - p_i = (4, 6)$  なので， $p_{i+1}$  は図 2 の領域 C にあるため， $v_i = (1, 1)$  になる．

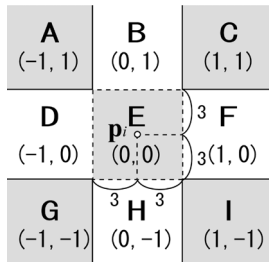


図 2: 移動方向  $v_i$  を決定する 9 つの領域

前式の U および V の定義に従って求めたモデル動作におけるマーカの位置および移動方向の時系列を  $U_0, V_0$  とし，ユーザ動作におけるマーカの位置および移動方向の時系列を  $U_1, V_1$  とする．マーカを複数使用する場合は，以下のように定義する．

$$\begin{aligned}
 U_{hm} &: \text{マーカ } m \text{ の位置の時系列} \\
 V_{hm} &: \text{マーカ } m \text{ の移動方向の時系列} \\
 0 \leq m < M (M: \text{マーカ数}) \\
 h &= \begin{cases} 0: \text{モデル動作} \\ 1: \text{ユーザ動作} \end{cases}
 \end{aligned}$$

## 2.2.2 DP マッチングによる動作の採点

### (1) DP マッチング

ユーザ動作とモデル動作の類似度を採点するために DP マッチング [7] を用いる．DP マッチングはパターンマッチング法の 1 つであり，2 つの文字列を比較し文字列間の類似度を求めることができる．また，DP マッチングには，類似度を求める過程で 2 つの文字列が全体的にできるだけ等しくなるように，モデルとなる文字列の一部に対して入力された文字列の一部の要素を伸縮して対応づける特徴がある．この特徴を利用して，モデル動作とユーザ動作におけるマーカの移動方向の時系列  $V_0, V_1$  に対して DP マッチングを適用すると， $V_1$  における各時点での移動方向を時間的に遅延あるいは先行させて  $V_0$  における各時点と対応づけを行う．これにより，フィードバック部で必要となる

モデル動作に対するユーザ動作の遅延あるいは先行などの情報を取得できる．ユーザに適切な動作指示を与えるためにこの情報を使うことができると考え，DP マッチングを採用した．

### (2) 動作の採点

$V_0$  と  $V_1$  に対して DP マッチングを適用すると，移動方向の推移に関するスコアを採点できる．ここで， $V_0$  の  $i$  番目の要素  $v_{0i} = (v_{0xi}, v_{0yi})$  と  $V_1$  の  $j$  番目の要素  $v_{1j} = (v_{1xj}, v_{1yj})$  の距離  $d(i, j)$  を以下のように定義する．

$$d(i, j) = \max \{ |v_{0xi} - v_{1xj}|, |v_{0yi} - v_{1yj}| \} \quad (1)$$

マーカが複数の場合は，マーカ  $m$  に対して式 (1) により  $d_m(i, j)$  を求め，その総和を距離  $d(i, j)$  と定義する．すなわち，

$$d(i, j) = \sum_{m=0}^{M-1} d_m(i, j)$$

である．DP マッチングの計算は以下のような 2 次元のマッチング行列  $g(i, j)$  を用いて行う．

$$g(i, j) = \min \begin{cases} g(i-1, j) + 2d(i, j) \\ g(i-1, j-1) + 4d(i, j) \\ g(i, j-1) + 2d(i, j) \end{cases}$$

$$\text{初期条件} \quad g(0, 0) = 0$$

$$g(i, 0) = g(0, j) = \infty$$

$$(1 \leq i \leq N-2, 1 \leq j \leq N-2)$$

上式により  $g(N-2, N-2)$  を計算する． $g(N-2, N-2)$  はモデル動作とユーザ動作の移動方向の類似度を表し， $g(N-2, N-2)$  の値が小さいほど移動方向はより類似していると判断できる．本システムでは， $g(N-2, N-2)$  の値をもとに移動方向のスコア  $e_{vec}$  を 100 点満点で算出する．

仮に  $V_0$  と  $V_1$  が一致したとするならば，上記のことから両者の移動方向は一致すると考えられる．しかし，ユーザ動作とモデル動作の移動方向が一致しても軌跡が異なる場合がある (図 3)．

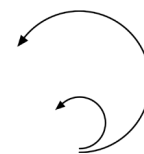


図 3: 移動方向は一致するが軌跡が異なる動作の軌跡

そこで，軌跡の違いを考慮した動作の採点も必要となる．DP マッチングによって得られた  $V_0$  と  $V_1$  の各時点での対応づけをそのままモデルと

ユーザのマーカ位置の時系列  $U_0, U_1$  にあてはめ、DP マッチングで対応づけた各時点での  $U_0$  と  $U_1$  のユークリッド距離の総和を求めることで、両者のマーカ位置についてずれの累積が求まり、軌跡の違いを表す量とすることができる。この軌跡の違いをもとに、軌跡の類似度に関するスコア  $e_{dis}$  を 100 点満点で算出する。

以上から、動作スコア  $e_m$  は動作の移動方向のスコア  $e_{vec}$  と軌跡のスコア  $e_{dis}$  をもとに算出される。具体的には式 (2) のようにこれら 2 つのスコアに重み  $w_{vec}, w_{dis}$  をつけて計算される。

$$e_m = w_{vec}e_{vec} + w_{dis}e_{dis} \quad (2)$$

ただし、 $w_{vec} + w_{dis} = 1.0$

$w_{vec}, w_{dis}$  を変えることで柔軟な動作採点が可能となる。例えば、モデルとユーザの体の大きさが異なる場合、 $w_{vec} = 1.0, w_{dis} = 0$  として動作の移動方向を重視した採点を行うなど、目的に合わせて適切に決定する。

### 2.2.3 ユーザ動作の判定

本節では、DP マッチングの結果をもとに、ユーザ動作がモデル動作に比べて遅延あるいは先行するなどを判定する方法について述べる。図 4 は DP マッチングにより求められたモデル動作とユーザ動作の対応づけを示している。図 4 の  $i$  軸はユーザ動作におけるマーカ移動方向の時系列  $V_1$  を表し、 $j$  軸はモデル動作におけるマーカ移動方向の時系列  $V_0$  を表す。

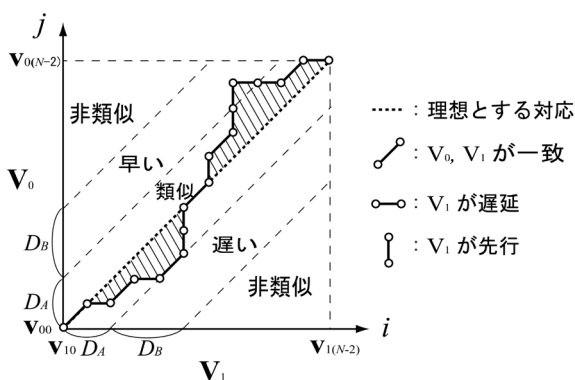


図 4: DP マッチングによる移動方向の対応づけ

モデル動作とユーザ動作の移動方向が完全に一致する場合、グラフは  $(V_{10}, V_{00})$  と  $(V_{1(N-2)}, V_{0(N-2)})$  を結ぶ直線 (以降、理想直線と呼ぶ) となる。ユーザ動作がモデル動作よりも遅延する場合、グラフは  $i$  軸正方向に推移し、逆にユーザ動作がモデル動作よりも先行する場合、グラフは  $j$

軸正方向に推移する。このような特徴から以下のことがわかる。

グラフの推移が平均して理想直線から上下に  $D_A$  以内に収まる場合、2.2.2 節より、移動方向が一致すると考えられる。グラフの推移が平均して理想直線から上下に  $D_A$  以上かつ  $D_A + D_B$  以内に収まる場合、グラフの推移の特徴から、モデル動作に対してユーザ動作の先行または遅延を判定できる。グラフがそれ以外の領域に推移する場合、動作はタイミングが合わず類似しないと判定する。また、経験により  $D_A = 0.2[s]$ 、 $D_B = 1.1[s]$  程度が適当であるとわかった。

### 2.3 音声採点部

振りの音声部分を評価、採点するためにマイクから入力される音声からリアルタイムに特徴量を抽出する。用いる特徴量は現時点では各フレーム内でマイクから入力される音声のパワーの平均値である。パワーを採用した理由は、振りのある決まった時点で手本と同様に発声ができているかどうかについて、パワーの特徴量で評価できると考えたためである。

音声評価、採点においては、特徴量が前フレームから閾値以上変化しているフレームに注目する。モデル音声においてこのようなフレームをすべて検出し、ユーザ音声においても同様の処理を行う。モデル音声とユーザ音声から検出された各フレームについて両者のパワーの変化がどの程度一致しているかを判定する。そこで、モデル音声において特徴量が閾値以上変化しているフレーム数を  $n$ 、モデル音声において特徴量が閾値以上変化しているフレームの中で、さらにユーザ音声とモデル音声の特徴量の変化が一致しているフレーム数を  $m$  とすると、ユーザ音声において特徴量の変化が検出されたフレームのうちモデル音声と特徴量の変化が一致している割合  $p$  が式 (3) のように求まる。

$$p = m/n \quad (3)$$

一致率  $p$  を 100 倍して音声のスコア  $e_v$  を 100 点満点で算出する。

### 2.4 振り採点部

振りのスコアは、動作と音声の各スコア  $e_m, e_v$  と両者を組み合わせたスコア  $e_c$  の 3 つのスコアをもとに算出される。具体的には、振りのスコア  $e_t$



は式 (4) のようにこれら 3 つのスコアに重み  $w_m$ ,  $w_v$ ,  $w_c$  をつけて計算される .

$$e_t = w_m e_m + w_v e_v + w_c e_c \quad (4)$$

$$\text{ただし } w_m + w_v + w_c = 1.0$$

$$e_c = (e_m + e_v)/2$$

重み  $w_m$ ,  $w_v$ ,  $w_c$  を変えることで様々な振りが評価できるようになる . 例えば , 動作のみの振りの場合や , 音声と動作が共に含まれる振りであっても , 動作のみを採点したい場合は  $w_m = 1.0$ ,  $w_v = 0$ ,  $w_c = 0$  とすればよい . 振りにおける音声のみを採点したい場合は  $w_m = 0$ ,  $w_v = 1.0$ ,  $w_c = 0$  , 動作と音声の両者ができないと意味がないような振りの場合は  $w_m = 0$ ,  $w_v = 0$ ,  $w_c = 1.0$  とする . また , 動作と音声を独立に評価したい場合は ,  $w_m$ ,  $w_v$  の値をそれぞれ 0.5 とする .

## 2.5 フィードバック部

フィードバック部には , 以下の機能がある .

- 振りを一通り終えた時に振りのスコアを表示し (図 5) , スコアに応じてエフェクト (効果音) を鳴らす
- リプレイ動画を再生するときに , ユーザに対する動作指示を表示する
- 振りを行っている最中に振りのスコアに応じてエフェクト (映像) を表示する

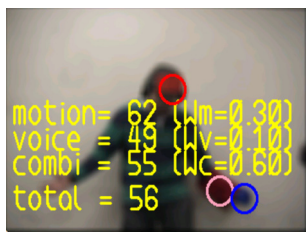


図 5: 振りに関するスコアの表示

以下 , リプレイ動画機能およびエフェクト機能について説明する .

### 2.5.1 リプレイ動画機能

リプレイ動画によりユーザが今行った振りについて , 2.2 節をもとに 6 種類の動作指示 (図 6,7,8) を 1 秒毎に表示する . ここで , リプレイ動画にはユーザおよびモデル動作のゴーストを重ねたものを用いて 1/2 倍速で再生する . これにより , ユーザは自分とモデルの動作を同時に見ることができ , 直観的にどこが誤りなのかを理解し易くなる . ユーザ動作が先行していれば「もっと速く」

と指示し赤く着色したゴーストを表示し , ユーザ動作が遅延していれば「もっとゆっくり」と指示し青く着色したゴーストを表示する (図 6) . モデル動作とユーザ動作の移動方向が一致する場合 , さらに位置について評価する . ユーザ動作の軌跡が小さいならば「もっと大きく」 , ユーザ動作の軌跡が大きいならば「もっと小さく」と指示する (図 7) . さらに , 両者の軌跡が一致するならば「素晴らしい」と褒める . 両者の移動方向が一致しなければ「もっと手本を見て」と指示しつつ緑に着色したゴーストを表示する (図 8) .

音声については , リプレイ動画の下部にモデル音声のパワー波形の推移をあらかじめ表示しておき , 前回行った振りのユーザ音声をリプレイ動画と同時に重ねて表示していく . これによりユーザは前回どのタイミングで発声し , モデル音声とどの程度ずれているのか理解できる (図 6,7,8) .

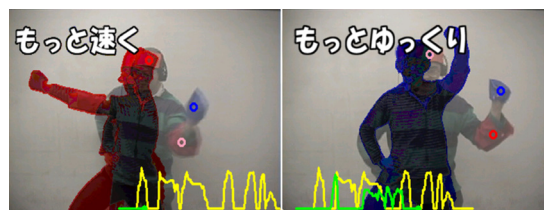


図 6: 「もっと速く」(左)と「もっとゆっくり」(右)

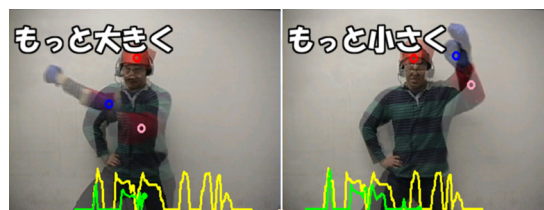


図 7: 「もっと大きく」(左)と「もっと小さく」(右)

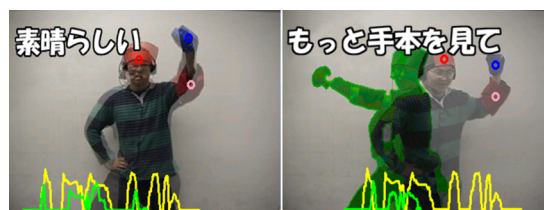


図 8: 「素晴らしい」(左)と「もっと手本を見て」(右)

### 2.5.2 エフェクト機能

振りのスコアに応じて次回の練習時にエフェクトを適用し , ユーザに振りの上達に対する達成感を与え , 練習における退屈感を解消する . エフェクトには , 練習の最中にリアルタイムに出力映像を加工する映像エフェクトと , 練習を終えて振りのスコアを表示するとき鳴らす効果音がある . 以下 , 本システムで用いる各エフェクトについて述

べる。

(1) 背景エフェクト (黒一色 / 綺麗な背景)

出力映像のユーザを除いた背景を切り替える映像エフェクトである。背景エフェクトは、出力映像から背景差分を用いてユーザ映像を切り出し、あらかじめ用意された綺麗な映像あるいは黒一色の映像にユーザ映像を上書きして実現する。

(2) ぼかしエフェクト

出力映像全体をぼかす映像エフェクトである。ぼかしエフェクトは、出力映像に平均値フィルタを適用して実現する。平均値フィルタとは、注目画素とその近隣の画素の平均画素値を注目画素に代入する手法である。

(3) 炎エフェクト

出力映像において、ユーザが装着するマーカ部分に炎を描画する映像エフェクトである。炎エフェクトは、炎の芯となる白い円を描画し、毎フレーム画像の下から上に向かって1画素ずつ平均値フィルタを適用して画素値を減少させてゆく。すると画素値の性質上、白が黄色になりさらに赤になるため、炎の芯を始点にして上に向かって黄色と赤の画素が流れていく様子を描画でき、まるで炎のような映像エフェクトが描画される。炎はある程度成長してから消えるため、ユーザがマーカを動かすと炎による軌跡ができる。

(4) 効果音エフェクト

振りのスコアに応じた効果音を出力する。スコアが高ければ明るい印象の効果音を鳴らし、スコアが低ければ暗い印象の効果音を鳴らす。

表 1: 振りのスコアと総合評価の対応

総合評価	振りのスコア
A	$70 \leq e_t \leq 100$
B	$60 \leq e_t < 70$
C	$40 \leq e_t < 60$
D	$00 \leq e_t < 40$

表 2: 映像エフェクトの切替

エフェクト \ 総合評価	A	B	C	D
炎	○			
綺麗な背景	○	○		
ぼかし			○	○
黒一色背景				○

式 (4) で得られた振りのスコア  $e_t$  により A から D の 4 段階で総合評価を決定する。振りのスコアと総合評価の対応を表 1 に示す。総合評価により、

表 2 のように映像エフェクトを切り替える。総合評価 A, B のときは、ユーザを誉めるということで、綺麗な背景を表示し、総合評価 A のときはさらにユーザのマーカ軌跡に炎を描画する映像エフェクトを加える (図 9)。総合評価 C のときは、まだまだという印象を与えるために出力映像全体をぼかす映像エフェクトを加える (図 10 左)。総合評価 D のときは、ユーザにまったく駄目であるという印象を与えるために、出力映像全体をぼかす映像エフェクトと黒一色の背景映像を表示する映像エフェクトを加える (図 10 右)。また、これらの映像エフェクトの他、総合評価 A, B のときは明るい印象の効果音を鳴らし、評価 C, D のときは暗い印象の効果音を鳴らす。



図 9: 評価 A のエフェクト (左) と評価 B のエフェクト (右)



図 10: 評価 C のエフェクト (左) と評価 D のエフェクト (右)

### 3 実験と評価

本システムにおける練習支援方法の有効性を検証する実験として、振りの練習を実際に被験者に行ってもらい、振りのスコアの推移を測定した。

#### 3.1 実験方法

被験者は大学生 20 名である。対象となる振りは、応援団経験者による応援の振りの一部 (約 5 秒) であり、掛け声を含む。被験者は対象の振りについて前知識がなく、本システムを用いて初めて振りの練習を行った。また、被験者は以下に示す 4 つのグループにそれぞれ 5 人ずつに分かれ、個々に振りの練習を 10 回行ってもらい、振りのスコアの推移を測定した。振り採点における重みづけは、式 (2) では  $w_{vec} = 0.75, w_{dis} = 0.25$  とし、式 (4) では  $w_m = 0.3, w_v = 0.1, w_c = 0.6$  とした。

#### グループ (1) 鏡像とビデオによる実験

鏡とビデオを使った練習方法である。ユーザは

手本となるモデル動作のビデオを見て振りを覚え、次にカメラの前でディスプレイに鏡像として表示される自分の振りを確認しながら練習する。

### グループ (2) ゴーストを用いた実験

従来研究 [2][3] のように、練習の最中のユーザが映っている映像にゴーストを重ねて表示することで、ユーザはゴーストに自分の姿を重ね、ゴーストに合わせて体を動かし、振りを練習する。

### グループ (3) 本システムを用いた実験

本システムを用いて練習する。

### グループ (4) ゴーストなしのリプレイ動画を用いた実験

グループ (3) と同様に本システムを用いて練習を行う。ただし、リプレイ動画ではユーザが今行った振りのみが再生され、ゴーストは表示されない。ユーザはそのリプレイ動画を見て振りの誤りを認識し、次の練習を行う。

## 3.2 実験結果と評価

実験結果を図 11 に示す。グラフは各グループについて被験者の振りのスコアの平均をとり、その推移を対数近似したものである。どのグループについても試行回数が増えるにつれてスコアが上昇していく傾向が見られた。

グループ (1) は、安定してスコアが上昇しているが、グループの中で一番低いスコアである。グループ (2) は、グループ (1) とほぼ同じ推移が見られるが、グループ (1) よりも高いスコアを安定して得ている。また、ゴーストを消しても振りが正確にできるかどうかを確認するために、ゴーストを用いた 10 回の練習後にゴーストを消して練習を 1 回行ったところ、平均でスコアが 6.0 下降した。このことから、ゴーストを用いた練習では、一定のスコアまでは早く到達できるが、回数を重ねてもグループ (1) とほとんど同じ程度までしか覚えられないことが確認できた。グループ (3) は、試行

回数の少ない内のスコアの上昇度が他のグループよりも高く、全体的に高いスコアを得ていることから、本システムが振りの練習に有効であることが確認できた。グループ (4) は、グループ (1), (2) よりもスコアは上昇していることからリプレイ動画を見ることによる練習効果が確認できた。ただし、グループ (3) よりも低いスコアに留まったことから、本システムにおけるリプレイ動画のゴーストの表示が学習支援に有効であると考えられる。

また、各グループに対して実験後に退屈感に関するアンケートに答えてもらった。ユーザはアンケート項目に対して 5 段階評価で降順に、非常にそう思う、そう思う、どちらでもない、そう思わない、まったくそう思わないと答えるものとした。表 3 の (1)~(4) はグループ番号であり、枠内の数字は評価値の平均である。炎エフェクトに関しては振りのスコアが 70 に達しなかった被験者は考慮しない。また、本システムを用いたグループ (3), (4) については映像エフェクトと効果音についての印象を聞いた。

グループ (1) では、モデル動作を見るときに自分の姿も同時に見たいという意見が多かった。また、鏡像とビデオを見て練習する方法はユーザが一方向的にシステムに振りを採点されるだけでつまらないという感想があった。グループ (2) では、ゴーストを目で追って追跡するためにどうしてもゴーストが先行してしまい体を合わせるのが難しいという感想があった。一方で、ゴーストを追跡する行為にはユーザの興味を引きつける効果があることがわかった。グループ (3) では、ゴーストを重ねるリプレイ動画を用いることで動作の誤りについて理解でき、次の練習時に参考になったという感想があった。逆に、リプレイ動画で動作と音声の誤りを同時に指摘されると両方を理解するのは難しいという感想もあった。グループ (4) では、リプレイ動画で自分の姿を見るだけではモデル動作との違いを具体的に把握できないので、リプレイ動画よりもモデル動作のビデオを重視して見る被験者が多かった。

## 4 おわりに

本研究では、個人で振りを練習する場合に、ユーザの動作と音声を評価、採点し、動作の誤りを補正する 6 種類の指示を行うことで練習を支援するシステムを提案、実装、評価した。評価のた

表 3: アンケート結果

アンケート項目 \ グループ番号	(1)	(2)	(3)	(4)
飽きずに練習を続けられた	3.0	4.2	4.0	4.0
炎のエフェクトが面白いと感じた			3.5	3.7
ぼかしエフェクトが面白いと感じた			3.2	3.4
背景が切り替わるエフェクトが面白いと感じた			3.0	3.6
効果音が面白いと感じた			4.4	4.3



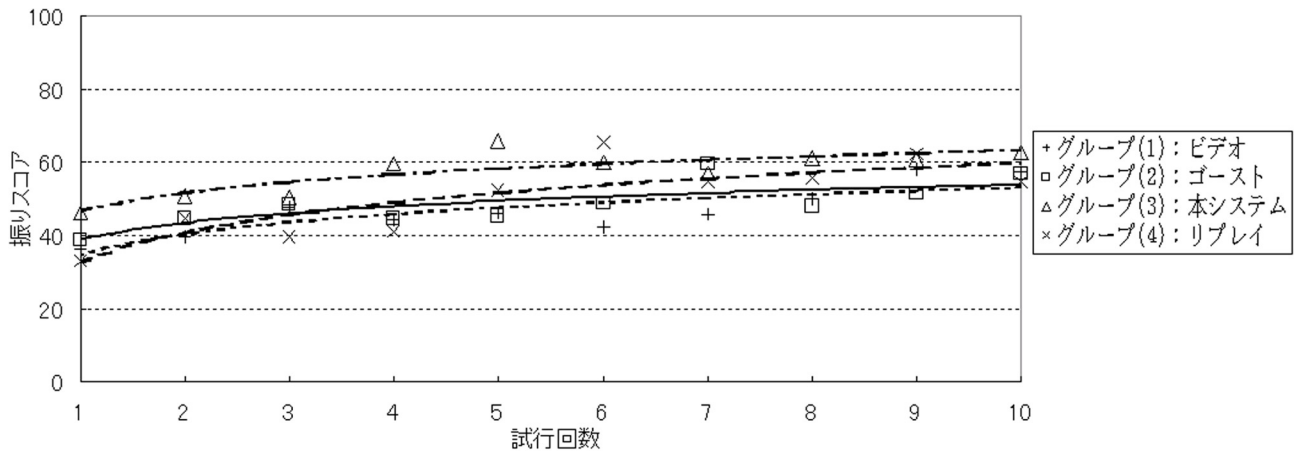


図 11: 各グループにおける振りのスコアの推移

めに、実際に被験者に振りの練習をしてもらい振りのスコアの推移を調べる実験とアンケート調査を行った。

本研究で提案する練習支援方法を用いると、ユーザは具体的に動作の誤りを把握でき、早い段階である程度の振りができるようになることがわかった。また、エフェクトについてはユーザの動きに反応する炎エフェクトなどの映像エフェクトを用いることにより、練習時に没入感を与えることができた。

実験により、グループ (2) のように振りの最中にゴーストを用いても到達できるスコアは本システムを用いる場合よりも低かったが、本システムよりも没入感が得られることがわかった。そこで、振りを覚えていない序盤はゴーストを用いて練習し、慣れてきたところでゴーストなしで練習していき、確認するときにはリプレイ動画にゴーストを用いる本手法へと切り替える練習方法を今後検討していきたい。

今後の課題として、リプレイ動画においては動作と音声の両方の誤りを同時に表示するとユーザの混乱を招く可能性があるため、両方をわかりやすく指摘する方法の検討がある。音声採点においては、現在は単純な特徴量を用いて掛け声など、発声のタイミングを重視する判定を行っているが、音韻情報やピッチ情報などより多くの特徴量を用いることで、より精度の高い採点ができるようになる必要がある。また、動作の奥行きを採点する必要がある振りの場合は、カメラを複数台使用するなどし、改善する必要があると考えている。

## 謝辞

画像処理に関し御指導いただいた、電気通信大学小池英樹助教授、東京農工大学中西泰人助教授、ソニーコンピュータサイエンス研究所高田哲司アソシエイトリサーチャーに深謝する。

## 参考文献

- [1] 総理府. 教育アンケート調査年鑑 2000 年版下, pp. 525-550, 創育社, 2001.
- [2] Ungeon Yang and Gerard Jounghyun Kim. "Just Follow Me": A VR-based Motion Training System, SIGGRAPH, 2001.
- [3] Ungeon Yang and Gerard Jounghyun Kim. "Just Follow Me": an Immersive VR-based Motion Training System, MIT Press, Vol. 11, No. 3, pp. 304-323, 2002.
- [4] 西本卓也, 園順一, 浅野令子, 高木治夫. 視覚障害者のためのタイピング練習ソフトの設計と評価, 電子情報通信学会技術研究報告, WIT 2001-11, pp. 7-12, 2001.
- [5] 吉本優子, 香川順子, 杉本圭優, 前迫考憲, 黒田卓, 森川治. 超鏡 (HyperMirror) 対話における教示行為に関する事例分析 「おりがみ」と「福笑い」を通して, 日本教育工学会誌, Vol.25-Suppl, pp17-22, 2001.
- [6] 福地健太郎, Ed Tannenbaum. "EffecTV": メガデモ技術のリアルタイムビデオエフェクトへの応用, エンタテインメントコンピューティング 2003 論文集, No. 1.
- [7] 中川聖一. パターン情報処理, 丸善, 1999.