論　文

ダンス動画への音声・視覚情報付与による低学年児童・幼児向けダンス習得支援システム

晴山　洋人[†]　　長谷川　忍[††]

A Dance Learning Support System for Lower-Grade and Preschool Children Using Audio and Visual Aids in Dance Videos

Hiroto HAREYAMA[†] and Shinobu HASEGAWA[††]

あらまし　本研究は, 見本動画を用いたダンス練習が一般化する一方, 児童・幼児が動画視聴のみで動作のタイミングや姿勢を正しく理解することが難しいという課題に着目した. 特にダンスの基本的な要素である, 姿勢を一瞬停止させる「止め」の動作に焦点を当て, 低学年児童・幼児を対象としたヒップホップダンス習得支援システムを開発し, その有効性を検証した. 本システムは, 音響・動画像解析により動画から「止め」のタイミングと姿勢を自動検出するコアエンジンと, 検出結果に基づきオノマトペ音声や視覚情報を付与する UI システムから構成される. 評価実験では, コアエンジンの最適手法を同定し, UI システムを用いて児童・幼児の練習効果を専門家が評価し, Wilcoxon の符号付順位和検定による統計的検討を行った. その結果, 短期練習では有意差は得られなかったものの, 女子のダンス経験者において「止め」の可視化が理解促進に寄与した可能性が示唆された. また, アンケートでは高い受容性が確認され, 特に視覚情報の有効性が顕著であった. 以上より, 本研究は従来研究で注目されなかった「止め」の自動検出技術を応用し, 児童・幼児向けダンス支援の新たな可能性を示すものである.

キーワード　ダンス練習, 自動検出, 児童・幼児, 音声付与, 視覚情報付与

1. はじめに

1.1 研究の背景

ヒップホップダンスの基本的なリズムの取り方は, 膝を屈伸させて沈み込む「ダウン」と膝を伸ばして体を引き上げる「アップ」である [1]. また, リズムをとりながら周期的に身体を「止め」る動作を循環的に行うのがヒップホップダンスの特徴である.

児童・幼児はヒップホップダンス（以下, ダンス）を基本的な動きから段階的に習得していく. 基本的な動きは, 姿勢を一瞬静止させる「止め」の連続で構成されており, 各ステップのカウントにおける「止め」の動きを通じて振付を学んでいく. [1] こうした動きを児童・幼児がダンスを習得する際には, ダンス教室などで指導者の動きを模倣することが一般的である. 内山はダンス学習の際に最もオーソドックスな方法は「模倣」であると述べている [2]. また飯野らは上級者の指導の下, 鏡で自分の姿を見ながら修正するか, DVD などの映像を見て真似るかのいずれかが主な練習方法であると主張している [3]. ダンス教室では, 指導者は児童・幼児の前に立ち, 鏡越しに後ろ向きで踊る. その結果, 児童・幼児は指導者の背面の動きと鏡に映った左右反転した動きを同時に観察しながら, 身体を動かし, 振付を学んでいく. また現代では, 指導者の動きを撮影し, その動画を家庭で確認しながら自主練習することも推奨されている.

しかし, 動画を視聴するだけでは, 児童・幼児にとって「止め」のタイミングやその姿勢を正しく理解することは難しいと考えられる. 実際のダンス指導の現場では「タン・タン」などのオノマトペ（擬音語）を用いて動きのタイミングを伝え「止め」の姿勢について指導者が説明を交えて指導し, それを児童・幼児が実践することで, 段階的に振付を習得している.

[†]＊
＊
[††] 北陸先端科学技術大学院大学　先端科学技術研究科
Japan Advanced Institute of Science and Technology
DOI:10.14923/transfunj.???????????

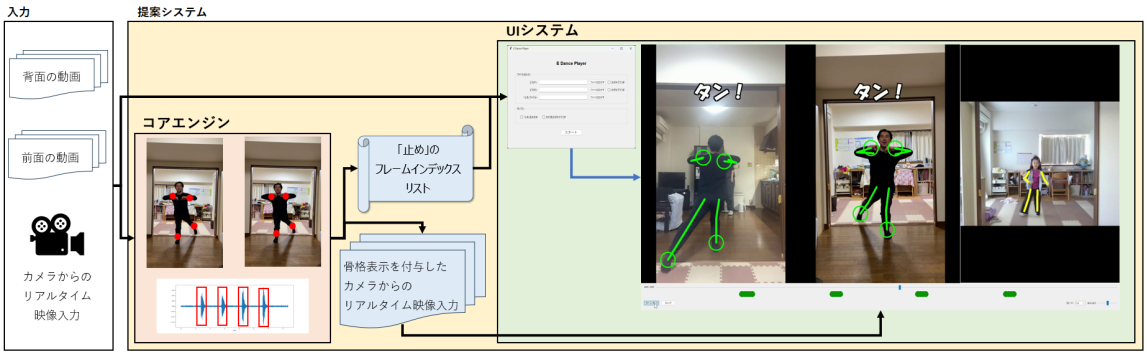


図 1: 提案手法のシステム概要

見本となる (ダンス指導者の) 動画からコアエンジンを通して「止め」のフレームインデックスリストを抽出する. また, 動画と同期撮影した背面動画を UI システムで入力する. 最後に, 見本の前面・背面の動画と, カメラからのリアルタイム入力にコアエンジンの処理を合わせた動画を UI システムで表示する.

1.2 研究の目的

そこで本研究では, ヒップホップダンスを対象とし, ヒップホップダンスにおける「止め」の動きに着目する. 動画上の「止め」の瞬間に音声・視覚情報を付与することで, 児童・幼児のダンス習得を支援することを目的とする. まず, 児童・幼児が見本動画から「止め」の動きを自動で検出できるシステムを開発する. 次に, 「止め」の動きに対して, 音声, 記号, 文字といった情報を付与することで, 児童・幼児にとってより分かりやすく振付を理解できるようにするシステムを構築する. さらに, 上記のシステムを用いて, 音声・視覚情報を付与した動画と, 付与していない動画を用いた場合で, 児童・幼児のダンス習得に差が生じるかを検証する比較実験を行う.

2. 関 連 研 究

2.1 教育的観点からのダンス習得に関する研究

高田は, 幼児期および児童期初期の発達段階に応じたダンス実践を報告しており, 幼児におけるダンス習得には「動作の分割（上半身と下半身を別々に動かすこと)」や「音楽に合わせて踊る時間の確保」が重要であるとしている [4]. 天野らは, 「言語のみ」「オノマトペのみ」「言語とオノマトペの併用」「カウントのみ」の 4 種類の指導方法が動作習得過程に与える影響について, 比較実験を通じて検討した. これらの指導方法はいずれも, 動画に音声を付加する形で提示された. その結果, 初めから「言語とオノマトペ」を併用して指導する場合, 学習者が情報過多になる可能性が示唆された. また, 初見の振り付けを学習する際には, まず全体の流れを把握し, その後に詳細な動作を段階的に習得するという学習プロセスが有効であることが示唆された [5].

2.2 システムによるダンス習得に関する研究

斎藤らは, 漫画風オノマトペをダンス動画に視覚的に付与することで, ダンス習得効果が向上することを示した [6]. また, Endo らは, ダンス動画から振り付けの短時間の動きを自動で分割する手法を提案しており, キーポイント間の速度変化を視覚特徴量として利用している [7]. さらに, Anderson らは, 自己学習を支援する AR ミラー型インターフェース「YouMove」を提案している [8]. 「YouMove」はユーザーの姿勢をリアルタイムに解析して見本との違いを可視化し, 動作の一時停止や繰り返し再生などの機能を備えている.

2.3 骨格検出に関する研究

Cao らによる OpenPose [9] は, 初めてマルチパーソン骨格検出をリアルタイムで実現したオープンソースのシステムである. ただし, リアルタイム処理については GPU を使用した場合に限り実現可能であることが報告されており, CPU のみでのリアルタイム実行は困難である. Bazarevsky らによる MediaPipe [10] は, Google が開発したオープンソースのマルチモーダル機械学習フレームワークである. リアルタイムな画像処理および機械学習パイプラインの構築を支援するプラットフォームで, 手や顔の検出, 姿勢推定などの高精度な機能を備えている. 少ない計算リソースでも一定の精度で実行できることが確認できたため, 本研究では MediaPipe を利用することとする.

3. 提案手法

3.1 アーキテクチャ

本研究は, 児童・幼児のダンス習得支援を目的として, 各ステップにおける「止め」の動作に着目し, 視覚および音声情報を付加することにより,「止め」の姿勢およびタイミングの理解を促進するシステムの提案を行うものである.

図 1 に提案システムのアーキテクチャ図を示す. 入力として見本となる (ダンス指導者の) 動画から「止め」のフレームインデックスリストを抽出する. また, 動画と同期撮影した背面動画を UI システムで入力する. 入力動画の「止め」のフレームに音声・視覚情報を付与し, カメラからのリアルタイム入力を合わせて UI システムで表示する. カメラからの表示には「止め」の抽出と同じアルゴリズムを用いて視覚情報を付与する.

提案システムは,「コアエンジン」と「ユーザーインターフェースシステム (以後 UI システム)」で構成される. コアエンジンでは, 入力されたダンス動画から音響情報および動画像情報を抽出する. 音響情報に関しては, 周期的な音のピークを検出し,「止め」のタイミングの候補フレームを抽出する. 一方, 動画像情報においては, 骨格推定によりダンス上級者の手首および足首のキーポイントを検出し, 各フレーム間における移動速度がゼロとなる箇所を「止め」の姿勢候補として抽出する. これら双方の候補が一致する動画フレームを「止め」の動作として確定する. UI システムでは, 検出された「止め」の動画フレームに対して, オノマトペによる音声情報および記号・文字による視覚情報を重ね合わせる. また, カメラからのリアルタイム入力にコアエンジンと同じアルゴリズムを用いて骨格情報を視覚的に付与する. これにより, 児童・幼児は視覚と聴覚の両面から「止め」のタイミングと姿勢を直感的に理解することが可能となる.

4. 実　　装

4.1 コアエンジン

コアエンジンの入力は動画 (.mp4) ファイルであり、出力は「止め」の動作を行っている動画フレームインデックスのリストである. コアエンジンでは, 音響情報で検出したフレームと動画像情報で検出したフレームの共通フレームを「止め」の動作を行っている動画フレームとして検出し, その動画フレームのインデックスのリストを出力する. また, ダンス経験者とのディスカッションの中で,「止ま」り始める部分についても「止め」の動作とするとよいとのアドバイスを受け, 共通フレームが連続 3 フレーム以下の場合は, 1 フレーム前のフレームも「止め」のフレームとした.

本研究での入力動画ファイルの条件を示す. 1. 入力動画ファイルは BPM(Beats Per Minute)=90 のダンス振り付けを撮影した動画である. 2. 入力動画ファイルの fps は 30.0 である. 3. 入力動画ファイルはメトロノーム音が鳴っている中で撮影した動画である. 4. メトロノーム音が鳴っているタイミングが「止め」のタイミングの候補となる. 5. メトロノーム音が動画内の音響情報において主要な要素を占めており, 他には足音などの微小な環境音がわずかに含まれるのみである.

4.1.1 音響情報による「止め」のタイミングの検出

音響情報による「止め」のタイミング検出手法として以下の手法を実装する.

入力動画の周期的な音響情報 (メトロノーム音) にて音の振幅がピーク (局所最大値) となる動画のフレーム番号を検出する. 音の振幅は動画内のフレーム数を N, 動画 1 フレーム単位での音の振幅 $x_i (i = 1, \ldots, N)$ とした時, 振幅の高さ h の条件 ($h = \mu + \sigma$, μ と σ は x_i の平均と分散である) を満たすフレームをピークとして検出する. ただし, 閾値で検出したフレームの前後 1 フレームもピークとする.

4.1.2 動画像情報による「止め」のタイミングの姿勢検出

動画像情報による「止め」のタイミング検出手法として以下の手法を実装する. これに先立ち, 前処理として各キーポイントの速度情報を算出する。まず, 骨格検出モデル MediaPipe [10] を用いて左右手首・足首のキーポイントを検出する. 次に, 検出したキーポイントの動画フレーム間速度を算出する. フレーム t で検出した i 番目のキーポイントの位置 (x_i, y_i) を $k_i(t) \in \mathbb{R}^2$, 速度 $v(t) \in \mathbb{R}^{(4\times2)}$ の i 番目の要素 $v_i(t) \in \mathbb{R}^2$ を以下の式で求める.

$$v_i(t) = |k_i(t) - k_i(t-1)| \tag{1}$$

上式で算出した速度 $v(t)$ が閾値以下のフレームを「止め」のタイミングの姿勢とした. 本研究では閾値の値を 10[pixel/frame] とする.

4.2 UI システム

UI システムの入力と出力は以下である. また, 図 2 の通り実装した.

入力: 同期撮影した見本のダンス動画ファイル (.mp4),

図 2: UI システム動作画面

(a) 背面動画, (b) 前面動画, (c) カメラ表示, (d) 動画の再生・停止, (e) 動画再生速度変更, (f) 音量調整, (g)「止め」タイミング印, (h)「止め」の際の骨格・図形表示, (i) オノマトペ表示, (j) カメラ動画像へのリアルタイム骨格表示

前面から撮影された動画ファイル, 背面から撮影された動画ファイル, コアエンジンで検出した「止め」の動作を行っている動画フレームインデックスのリスト, カメラからの動画像 (リアルタイム表示)(図 2(c))

出力: 入力された動画ファイルに音声・視覚情報を付与した動画 (図 2 (a), (b)), カメラからの動画像に視覚情報を付与したリアルタイム表示 (図 2 (j))

上記の”音声情報の付与”とは「止め」のタイミングでオノマトペの音声情報を付与することである.

上記の”視覚情報の付与”とは次に示すことを行うことである. 1. 左右の肩から手首, 腰から足首までに骨格表示を行う. 骨格表示は「止め」の姿勢では緑色になり, それ以外は黄色になる. (図 2 (h)), 2.「止め」の姿勢の際に左右手首・足首に丸の図形付与 (図 2(h)), 3.「止め」の姿勢の際に漫画風オノマトペの文字を付与 (図 2 (i)), 4. シークバーの「止め」のタイミングに緑の印を付与 (図 2 (g))

5. 実験・評価

5.1 コアエンジンの実験・評価

コアエンジンで適切に「止め」を検出できるかを確認することを目的に, 以下の実験・評価を行った.

5.1.1 コアエンジンの実験

1. 評価用動画を用意する. 本実験では 5 つの動画を評価に用いた. 2. ダンス経験者監修のもと「止め」のフレームにアノテーションを行った. 具体的には各動画フレームを一枚ずつ画像に分割し,「止め」のフレームだと考える動画フレームインデックスのリストを作成した. 3. コアエンジンで「止め」のフレームを推定した. 4. 推定した「止め」のフレームとアノテーションした「止め」のフレームが合致するかダイスインデックス (Sørensen-Dice coefficient) により評価した.

表 1: コアエンジン評価 音響情報による「止め」の検出手法比較

	Sample1	Sample2	Sample3	Sample4	Sample5
GT	12	21	18	26	21
cnt	34	32	26	43	28
TP	9	8	10	6	14
FP	25	24	16	37	14
FN	3	13	8	20	7
TN	175	167	178	149	177
Acc	0.868	0.825	0.887	0.731	0.901
Recall	0.75	0.381	0.556	0.231	0.667
Precision	0.265	0.25	0.385	0.14	0.5
Dice	0.391	0.302	0.455	0.174	0.571

5.1.2 コアエンジンの評価

上記により、ダイスインデックスを用いた音響情報による「止め」の検出評価と, 動画像情報による「止め」の検出評価をそれぞれ行った. また, ダイスインデックスが高い事例と低い事例に関する考察も行った.

表の文言整理: ここで, 表の文言は次の通りである. GT: アノテーションしたフレームの数. cnt: 検出したフレーム数. TP: 検出したフレームと GT が合致した数. FP: 検出したフレームと GT が合致しなかった数. FN: 検出しなかったフレームが GT であった数. TN: 検出しなかったフレームが GT でなかった数. Acc: 精度. Recall: 再現率. Precision: 適合率. Dice: ダイスインデックス.

コアエンジン評価の総括: 音響情報及び動画像情報による検出では, 各動画でダイスインデックスの値が 0.174 ~ 0.571 となり, また GT の合計数と TP 合計数から「止め」のアノテーションの約 1/2 を検出できたことが示され, コアエンジンの課題に寄与できたと考える. しかし, 以下考察の通り,「止め」を検出しにくい動画への対応が今後の課題である.

ダイスインデックスが高い事例と低い事例に関する考察: 今回の実験では, 動画 4 がどの手法も総じてダイスインデックスの値が低く, 動画 5 のダイスインデックスの値が最も高かった. 動画 4 のダンスは「止め」の動画の繰り返しでありつつも, やや流れるような動きであったため,「止め」を検出しにくかったと考えられる. また, 動画 4 のダンスは身体を大きく使う振付になっており, 反動をつけるために次の動作への予備動作が比較的大きくなったと考えられる. 対して, 動画 5

表 2: グループ別ダンス振付及び音声・視覚情報付与有無比較表

グループ	1 回目	2 回目
A	ダンス a: 音声・視覚 有	ダンス b: 音声・視覚 無
B	ダンス a: 音声・視覚 無	ダンス b: 音声・視覚 有
C	ダンス b: 音声・視覚 有	ダンス a: 音声・視覚 無
D	ダンス b: 音声・視覚 無	ダンス a: 音声・視覚 有

では 1 拍 1 拍を「止め」, 身体を使う範囲が比較的狭く予備動作も少ないことから, 検出がしやすかったと考えられる. 同様の考察は [7] でもなされており, 拍の動きにアクセントが来る振りの場合は [7] の論文で議論されている動画分割が行いやすく, 反対に柔らかく流れるような振付に対しては分割が難しかったと述べられている.

5.2 UI システムの実験・評価

UI システムを用いて児童・幼児にダンス練習を実施してもらい, 音声・視覚情報の有無でダンスの動きとリズムの習得度に変化があるか評価した. また, 音声・視覚情報の付与がダンス習得に有効に働いた群はどのような背景属性を持つか分析を行った. さらにアンケートにおいて音声・視覚情報の付与が児童・幼児にとって役に立ったか主観的な評価を行った.

5.2.1 UI システムの実験

実験参加者の属性は次の通りである. 男女人数：男 4 人, 女 12 人 (計 16 人), 年齢別任数：6 歳 9 人, 8 歳 6 人, 9 歳 1 人, 平均 6.917 歳, 標準偏差 1.165, ダンス歴有無：歴有 6 人, 歴無 10 人, 平均 0.396 年, 標準偏差 0.887　以下の通り実験参加者をグループに分け実験準備を行った (表 2).

1. 実験参加者を 4 グループ (A, B, C, D) に分ける.
2. ダンスの振付を 2 つ (ダンス a, ダンス b) 用意する.
3. 1 つのダンスの振付について, 音声・視覚情報を付与した動画 (付与有) で練習するグループと音声・視覚情報を付与しない動画 (付与無) で練習するグループに分ける.

以下の通り実験を行った. 実験参加者一人ずつ表 2 のグループ分けの通りの順番で練習を行った.

1. 練習するダンスを動画で 2 回確認する.
2. UI システムでの練習前にダンス振付を行い, それを撮影する.
3. UI システムを使用して 5 分間ダンス練習を行う.
4. UI システムでの練習後にダンス振付を行い, それを撮影する.

上記を 1 回目, 2 回目のダンスで行い, その後児童・幼児にはアンケートに回答してもらった. また, 練習前後で撮影したダンスを指導者に確認し, ダンスの動きとリズムについて評価を行った. 指導者はダンス歴 22 年, 指導歴 16 年の X 氏とダンス歴 13 年指導歴 6 か月の Y 氏に依頼した.

＜ *Wilcoxon* の符号付順位和検定＞

アンケート結果を集計し, 音声・視覚情報の有無により, 以下 4 つの項目について検定を行った. X 氏 Y 氏の評価については平均値を用いた.

1. X 氏 Y 氏が練習前後で評価した「動きがよくできたと思いますか？」の値の差の平均値
2. 児童・幼児が評価した「動きがよくできたと思いますか？」の値
3. X 氏 Y 氏が練習前後で評価した「リズムはよくできたと思いますか？」の値の差の平均値
4. 児童・幼児が評価した「リズムはよくできたと思いますか？」の値

上記については, 次の略称を以後使用する. 「1. X 氏 Y 氏 A(平均)」, 「2. 児童・幼児 A」, 「3. X 氏 Y 氏 R(平均)」, 「4. 児童・幼児 R」.

音声・視覚情報の有無による 2 群間で母集団の中央値に差があるかを検定するため, ノンパラメトリック手法である Wilcoxon の符号付順位和検定 (Wilcoxon signed-rank test) を実施した. 本検定はデータが正規分布に従わない場合でも有効であり, 本実験のような状況にも適している. 具体的な手順としては, 各ペアの差 $d_i = x_i - y_i (i = 1, \ldots, N)$ を計算し, 差が 0 でないものを抽出した. その際, 差が 0 のデータについては有効サンプル数 (N) から除外した. その後, 絶対値 $|d_i|$ に対して昇順に順位 (rank) を付け, 元の符号 (正負) を順位に戻した. 正の符号に対応する順位の総和 W^+ および負の符号に対応する順位の総和 W^- を算出した. 検定統計量 W は, これらのうちの小さい方 ($W = \min(W^+, W^-)$) を採用し, これを用いて「中央値に差がない」とする帰無仮説の検定を行った. 検定は両側検定で有意水準 $p = 0.05 or 0.01$ にて行った. 統計的有意差があると判断する統計表は Scipy [11] の scipy.stats.wilcoxon を用いて作成した.

音声・視覚情報の付与が有効に働いた群の背景分析: また, 今回実験したデータに対して解析を行った. 解析は, 被験者の背景属性および児童・幼児の自己評価項目について, 音声・視覚情報の付与が有効であった群 (Snd_Vis-oriented), 音声・視覚情報の付与のない方

が有効であった群 (Non-Snd_Vis-oriented), および両者に差が見られなかった群 (Balanced) の 3 カテゴリに分類し, それぞれの特徴を 5 段階スケールのレーダーチャートにより可視化した. まず, 音声・視覚情報の付与によるダンス習得効果を示す 4 つの評価指標 (音声・視覚情報有無それぞれの「音声・視覚情報有 X 氏 Y 氏 A(平均)」, 「音声・視覚情報有 X 氏 Y 氏 R(平均)」, 「音声・視覚情報無 X 氏 Y 氏 A(平均)」, 「音声・視覚情報無 X 氏 Y 氏 R(平均)」の 4 項目) に基づき, 各被験者に音声・視覚情報有りの総合得点 SV_i と音声・視覚情報無しの総合得点 NSV_i を以下の式により算出した ($i = 1, \ldots, 16$)：

$$SV_i = \sum_{j}^{4} sv_{ij}, \; NSV_i = \sum_{j}^{4} nsv_{ij} \tag{2}$$

ここで, sv_{ij} および nsv_{ij} は, それぞれ音声・視覚情報有無の各指標である. 次に, 音声・視覚情報有無の得点差 $D_i = SV_i - NSV_i$ に基づき, 以下の基準カテゴリを付与した：

- $D_i \geq 0.5$: Snd_Vis-oriented 群
- $D_i \leq 0.5$: Non-Snd_Vis-oriented 群
- 上記以外 : Balanced 群

このカテゴリ分類に基づき, 以下の 7 項目を対象に平均値を算出した：

背景属性：性別 (gender), 年齢 (age), ダンス歴 (dance history)

児童・幼児自己評価項目：

音声・視覚情報有_動き (snd_vis_Act_child),
音声・視覚情報有_リズム (snd_vis_Rhy_child),
音声・視覚情報無_動き (Non-snd_vis_Act_child),
音声・視覚情報無_リズム (Non-snd_vis_Rhy_child)

上記のうち, 性別・年齢・ダンス歴の 3 項目は尺度が異なるため, 5 段階にスケーリングした. これにより, すべての指標を 5 段階スケールで視覚的に比較可能な形式に統一した. 最終的にカテゴリごとの平均ベクトルをレーダーチャート上にプロットし, 各群の傾向を視覚化した.

アンケートによる音声・視覚情報付与手法別の主観的評価分析: さらに, 音声・視覚情報の付与が児童・幼児に役立ったか主観的な評価を行った.

5.2.2 UI システムの評価

＜ *Wilcoxon* の符号付順位和検定＞

音声・視覚情報の有無による Wilcoxon の符号付順位和検定では, 4 項目すべてにおいて有意差はみられなかった. すなわち, いずれの評価においても, 音声・視覚情報の付与による学習効果の違いは統計的に有意な変化を示さなかった. この結果は, 音声・視覚情報の有無が主観的な動作評価やリズム評価に与える影響が限定的である可能性を示唆している. 特に, 児童・幼児自身の評価 (項目 2 および 4) においても顕著な差が認められなかったことから, 音声・視覚情報の付与が学習者自身の動作感覚の変化に直結しない可能性がある. また, X 氏および Y 氏による評価においても一貫して非有意であったことは, 観察者による外的評価においても同様の傾向が見られることを意味する.

これらの結果は, 今回の実験設定 (回数) では音声・視覚情報の提示が必ずしも学習成果の向上につながるとは限らないこと, あるいは提示方法やタイミングなど, 他の要因が効果に影響している可能性を示唆する.

表 3: 音声・視覚情報の有無による Wilcoxon の符号付順位和検定結果

項目	N	W	p=0.05	p=0.01
1. X 氏 Y 氏 A(平均)	12	37.5	-	-
2. 児童・幼児 A	14	36.5	-	-
3. X 氏 Y 氏 R(平均)	12	28	-	-
4. 児童・幼児 R	9	20	-	-

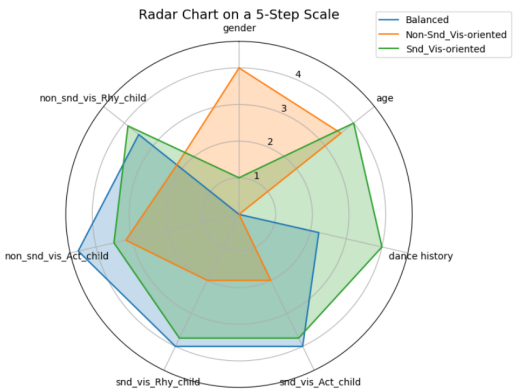


図 3: 音声・視覚情報有無による有効群ごとの背景情報レーダーチャート

＜音声・視覚情報の付与が有効な群の背景分析＞

次に, 音声・視覚情報の付与が有効であった群 (Snd_Vis-oriented, $n = 8$), 音声・視覚情報の付与のない方が有効であった群 (Non-Snd_Vis-oriented, $n = 6$), および両者に差が見られなかった群 (Balanced, $n = 2$) の 3 群について, 各群の性別・年齢・ダンス歴と自己評価項目との関連性を分析した (図 3, 表 4). ここで, 表

表 4: 音声・視覚情報有無による有効群ごとの背景情報表

	人数	1 女, 2 男	年齢	ダンス歴 (年)	snd_vis		Non-snd_vis	
					Act_child	Rhy_child	Act_child	Rhy_child
Snd_Vis oriented	8	1.13	7.13	1.34	3.75	3.75	3.50	3.88
Non-Snd_Vis oriented	6	1.50	7.00	0.00	2.00	2.00	3.17	2.17
Balanced	2	1.00	6.00	0.75	4.00	4.00	4.50	3.50

4 は図 3 の数値を表にまとめたものである. また, 表の性別・年齢・ダンス歴は正規化前の実際の平均値を使用している.

■ **Snd_Vis-oriented 群の特徴**

Snd_Vis-oriented 群 (n = 8) は, 性別平均が 1.13(≒女子中心), 年齢 7.13 歳, ダンス歴 1.34 年と, 他群と比較してダンス経験が長く, 年齢も高めであった. また, 自己評価においても「snd_vis_Act_child」と「snd_vis_Rhy_child」の得点がいずれも 3.75 と高く, 音声・視覚情報提示に対する感受性が高い傾向が見られた. 加えて, 自己評価スコア「non_snd_vis_Act_child」(3.50),「non_snd_vis_Rhy_child」(3.88) も一定以上であり, 全体的に自己評価の高い児童・幼児で構成されていると解釈できる.

■ **Non-Snd_Vis-oriented 群の特徴**

Non-Snd_Vis-oriented 群 (n = 6) は, 性別平均が 1.50(≒男子中心), 年齢 7.00 歳, ダンス歴 0.00 年であり, 未経験の男子児童が主に該当した. 自己評価において,「non-snd_vis_Act_child」は 3.17 と一定の高さを示した一方,「snd_vis_Act_child」と「snd_vis_Rhy_child」はともに 2.00 と低く, 音声・視覚情報提示による効果が出にくい層といえる.

■ **Balanced 群の特徴**

Balanced 群 (n = 2) は, 性別平均 1.00(女子のみ), 年齢 6.00 歳, ダンス歴 0.75 年と, 年齢・経験ともに中間的な位置にある. 自己評価スコアはいずれも高く,「snd_vis_Act_child」と「snd_vis_Rhy_child」は 4.00, また,「non_snd_vis_Act_child」は 4.50,「non_snd_vis_Rhy_child」は 3.50 と高水準にあった. これは, 音声・視覚情報付与いかんによらず, 柔軟に適応できる学習者像を示していると考えられる.

これらの結果は, 児童・幼児において, 性別やダンス経験年数によって音声・視覚情報の有無が学習効果に与える影響が異なる可能性を示唆している. 特に以下のようなことが考えられる.

- ダンス経験のある女子児童には動画への音声・視覚情報の付与が有効
- 未経験かつ男子児童には, 音声・視覚情報の付与の効果は限定的
- 音声・視覚情報の付与いかんに関わらず高得点を示す児童には状況に応じたダンス習得が望ましい.

表 5: 児童・幼児による音声・視覚情報付与手法の主観的評価

	Ave	StDev	Max	Min	Median
見本動画	3.938	0.937	5	3	4
カメラ表示	4.063	1.128	5	2	4.5
オノマトペの音	3.625	1.557	5	1	4
システム	4.125	0.953	5	3	4.5

また, 児童・幼児に「システムをもっとよくするためにこうした方がいいと思うところはありますか？」とアンケートしたところ, Non-Snd_Vis-oriented 群の児童・幼児から,「カメラ表示の記号の○と棒が混乱してしまった.」,「ダンスのことがわからない.」という意見があった. さらに, Balanced 群の児童・幼児からは「先生の動きにもっとあわせられるシステムだとよかった.」との意見があった. 考察として, ダンス未経験の児童・幼児はどこに意識を集中するかのイメージが難しく, 本提案システムはある程度ダンス経験がある方が有効である可能性がある. ＜アンケートによる音声・視覚情報付与手法別の主観的評価分析＞

さらに, 音声・視覚情報の付与が児童・幼児に役立ったかの主観的な評価を表 5 にまとめる. ここで, 表の文言は以下質問項目の略称である.

- 見本動画 : 見本動画の文字や記号はダンス練習の役に立ちましたか？
- カメラ表示 : カメラ表示の記号はダンス練習の役に立ちましたか？
- オノマトペの音 : オノマトペの音 (タン) はダンス練習の役に立ちましたか？
- システム : またシステムを使いたいと思いますか？

また, それぞれについて平均値 (Ave), 標準偏差 (StDev), 最大値 (Max), 最小値 (Min), 中央値 (Median) を算出

した．平均値に着目すると，「システム」が最も高く (4.125), 次いで「カメラ表示」(4.063),「見本動画」(3.938),「オノマトペの音」(3.625) の順となっており, 視覚的補助 (カメラ表示, 見本動画) に対する評価が音声的補助 (オノマトペ) よりも高い傾向が見られた．一方で, 評価のばらつきを示す標準偏差に着目すると,「オノマトペの音」が最も大きく (1.557), 児童・幼児間での評価の個人差が顕著であることが示唆される．最小値も 1.000 と, 他の項目に比べて顕著に低い．このことから, オノマトペによる提示は一部の児童・幼児にとっては理解や需要が難しい可能性がある．一方で「見本動画」の標準偏差は 1.0 未満であり,「カメラ表示」も「オノマトペの音」よりも標準偏差が低いことから, 視覚的補助は比較的安定した評価が得られている．また,「システム」は標準偏差が 0.953 であり, 児童・幼児には満足度が高い結果となったため, 練習の習慣化にも使用できる可能性が示唆された．

以上の結果から, 視覚的な情報提示 (「見本動画」,「カメラ表示」) は児童・幼児に対して有効であり, かつ評価のばらつきが小さいことから一貫した学習支援手法として有望であると考えられる．一方で, オノマトペの音声提示については, 平均値が 4.0 に近い水準を示しながらもばらつきが大きく, 個別の特性や学習スタイルに応じた柔軟な運用が求められる手法であるといえる．

6. おわりに

6.1 まとめ

本研究では, ヒップホップダンスにおける「止め」の動きに着目し, 動画上の「止め」の瞬間に音声・視覚情報を付与することで, 児童・幼児のダンス習得を支援するシステムを提案した．提案手法は大きくコアエンジンと UI システムとに大別され, コアエンジンでは動画を入力として音響情報と動画像情報により「止め」の動作を行っている動画フレームを検出し, UI システムでは検出した動画フレームに音声・視覚情報を付与することができた．UI システムを使用した児童・幼児に対する比較実験では, 音声・視覚情報を付与してダンス練習を行った群と音声・視覚情報を付与しないでダンス練習を行った群で統計的有意差が認められるか検定を行った．音声・視覚情報の有無で統計的な有意差はみられなかったが, 音声・視覚情報の付与を行うことがダンス習得に有効な児童・幼児の背景属性を分析することができた．また, 児童・幼児のシステムへの満足度が高く, 練習に不可欠な継続性の可能性を見いだせたこと, 視覚情報の付与が学習支援として有効であること, 音声情報の付与は児童・幼児の特性を考慮して選択的に使用する必要性があることなどの示唆が得られた．

謝辞　実験データの提供および評価にご協力いただいた DANCE STUDIO NEST の先生方, ならびに実験データ収集にご協力くださった児童・幼児とその保護者の皆様にも, この場を借りて深く御礼申し上げます．

文　献

[1] 柳瀬慶子，ヤナセケイコ，“表現運動・ダンス学習におけるリズム系ダンスの「リズムの特徴を捉えて踊る」ということに関する考察: 小学校「サンバのリズム」と中学校・高等学校「ヒップホップのリズム」に着目して，” PhD thesis, Tokoha University, Tokoha University Junior College Repository, 2024.

[2] 内山須美子，ウチヤマスミコ，“ストリートダンスのステップを用いた定形型ステップ学習の教育的意義と課題，” 白鴎大学教育学部論集，vol.10，no.1，pp.95–126，2016.

[3] 飯野友里恵，森谷友昭，高橋時市郎，“ストリートダンス動作の分析とダンス指導への応用 (映像表現フォーラム)，” 映像情報メディア学会技術報告 35.14 一般社団法人 映像情報メディア学会，pp.49–52 2011.

[4] 高田康史，“幼児・児童にもできる簡単ヒップホップダンスに関する実践報告―ipu わくわくリズムダンスの実践を通して―，” PhD thesis，International Pacific University，2015.

[5] 天野海都，三浦健，椙ちか子，“ダンス動画を用いたストリートダンス指導における伝達方法の違いが動作習得過程に及ぼす影響，” スポーツパフォーマンス研究，vol.15，pp.176–185，2023.

[6] 斎藤光，徳久弘樹，中村聡史，小松孝徳，“ダンス動画へのオノマトペ付与によるダンス習得促進手法，” Technical report，情報処理学会，2020.

[7] K. Endo, S. Tsuchida, T. Fukusato, and T. Igarashi, “Automatic dance video segmentation for understanding choreography,” Proceedings of the 9th International Conference on Movement and Computing, pp.1–9, 2024.

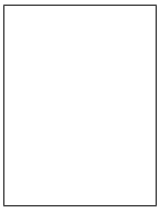
[8] F. Anderson, T. Grossman, J. Matejka, and G. Fitzmaurice, “Youmove: enhancing movement training with an augmented reality mirror,” Proceedings of the 26th annual ACM symposium on User interface software and technology, pp.311–320, 2013.

[9] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, “Openpose: Realtime multi-person 2d pose estimation using part affinity fields,” IEEE transactions on pattern analysis and machine intelligence, vol.43, no.1, pp.172–186, 2019.

[10] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, “Blazepose: On-device real-time body pose tracking,” arXiv preprint arXiv:2006.10204, pp.••–••, 2020.

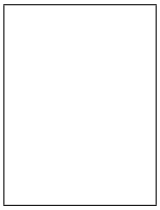
[11] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al., “Scipy 1.0: fundamental algorithms for scientific computing in python,” Nature methods, vol.17, no.3, pp.261–272, 2020.

（xxxx 年 xx 月 xx 日受付）



晴山　洋人

2012 年 首都大学東京都市環境学部都市環境学科地理環境コース卒 (学士 (理学))
2025 年 北陸先端科学技術大学院大学先端科学技術研究科先端科学技術専攻修了 (修士 (情報科学)).
現在，一般企業に従事



長谷川　忍

*

Abstract This study proposes a hip-hop dance learning support system for young children, addressing difficulties in understanding timing and posture from videos alone. Focusing on the fundamental “stops” pose, the system combines automatic stops detection using audio-visual analysis with a UI providing onomatopoeic cues and visual feedback. Expert evaluations and Wilcoxon tests showed no overall short-term significance but suggested benefits for experienced girls, with strong user acceptance. These findings indicate the potential of stops detection for early childhood dance education.

Key words dance practice, automatic detection, children, audio support, visual support