＜紙芝居上演における共感の可視化の試み＞

研究年度　令和 6 年度
研究期間令和 6 度～令和 6 年度
研究代表者名　前村　葉子
共同研究者名　柳田　多聞

1.はじめに

紙芝居は日本発祥のストーリーテリングの手法であり，保育園や高齢者福祉施設におけるレクレーションなどで使われることが多い．紙芝居上演の場は，舞台，演技者，観客から構成され，3 者間の相互作用により共感がうまれ上演の場の雰囲気が醸成され参加者に癒しや気分転換をあたえる．共感は人々の社会生活の基盤となるものであり，コミュニケーションスキルと密接に関連している．優れた演者は観客の表情や態度から退屈していないか，楽しんでいるかを読み取り，話し方やジェスチャなどを臨機応変に調整する．このような観客の反応に共感し適応する話し方を技術的に支援・学習できるようなシステムの創発が求められている[1,2]．また，音声，視線，ジェスチャ，表情などのマルチモーダル信号をテキストや感情と統合的に学習することで，対人コミュニケーションを理解・支援する人工知能技術が注目されており[3,4,5]，紙芝居上演の形式は簡潔であり演者の演技行動は構造化と親和性が高いことから，演者の演技行動データを収集・分析することで共感反応の定量的な評価に寄与することが期待できる．そこで，本研究は，紙芝居における物語進行時の演者と観客行動を自動取得し定量的に可視化することを目的として，今年度は，紙芝居ストーリーテリングタイムラインの進行指標となる紙芝居画面が更新されるタイミングの自動検出手法，および，頭部姿勢検出技術を用いた演者のアイコンタクト動作および観客の注目状態推定手法を提案した．本提案にもとづく可視化手法をもちいることで紙芝居のみならずクラスルームのような話者と複数の視聴者からなる場における物語を介したアイコンタクト動作と応答をはじめとするコミュニケーション過程の客観的データ収集と分析の基盤の構築の一助になるものと考える．

2.紙芝居シーン変更点の検出

2.1 提案手法の概要

本研究では，紙芝居ストーリーテリングタイムラインの進行指標となる紙芝居画面の更新タイミング，これは現画面が次画面に切り替わるタイミング(以降,「シーン変更点」と呼ぶ)の自動検出法を提案した．提案手法の概要を図 1 に示す．

はじめに演者映像のフレームにおいて紙芝居の舞台領域を検出する．ここでは一般物体検出モデル YOLO[6]を基盤とし，紙芝居舞台領域検出に特化するためファインチューニングを実施した．つぎに，舞台にセットされた画面が変更されるタイミング検出には，密なオプティカルフローから水平方向移動速度を算出し閾値ベースの異常検出による信号処理手法によりシーン変更点を検出した．舞台領域検出タスク用の学習データおよびシーン変更点検出のための実験データとして，紙芝居抜き差し動作の専用データセットを構築し，学習および信号処理パラメータの最適化と検出精度評価をおこなった．

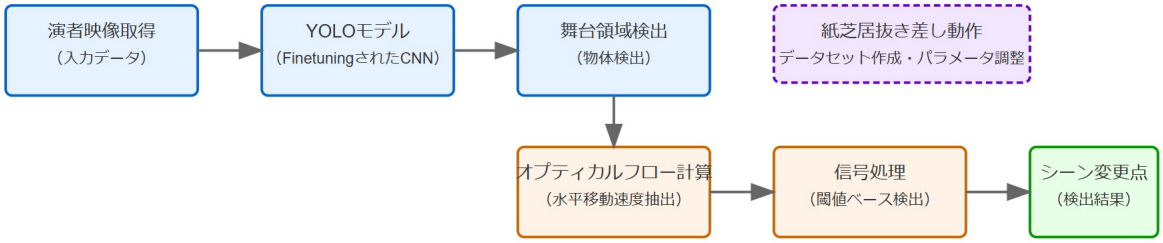

図 1: 紙芝居シーン変更点の検出の概要

2.2.データセット

紙芝居シーン変更点検出処理パラメータ最適化のため，音声資源コンソーシアム（SRC）提供の「日本語単一話者オーディオブック・紙芝居朗読音声コーパス（J-KAC）[7]」から入手可能な 15 作品を使用して演者データセットを構築した．研究室内でビデオカメラ(Panasonic HC-X2000（60fps，1920×1080 解像度）を三脚で固定して撮影した．演者は 2 名で重複なく作品を担当し，発話は行わず，画面引き抜き動作のみを含む簡略化された演技を行った．各作品は 12 シーンで構成され，シーン変更点が GrandTruth としてアノテーションされた．また，舞台領域検出のための基盤モデル FT 用学習データとして，本データセットから抽出したフレーム画像のサブセットを使用した．

2.3.舞台領域の検出

事前学習モデルには YOLOv8n を用いた．ファインチューニング用データセットとして，演者映像から抽出した 214 フレームに舞台の矩形領域をアノテーションして作成した．データセットは学習用と検証用に約 4:1 の比率で分割した．図 2 に学習曲線を示す．モデルのバックボーン，ネック，ヘッドの全レイヤーにおいて適切な収束が確認された．評価指標である Precision（適合率），Recall（再現率）および mAP(mean Average Precision)値から，検出精度と汎化性能の両面で良好な結果が得られた．

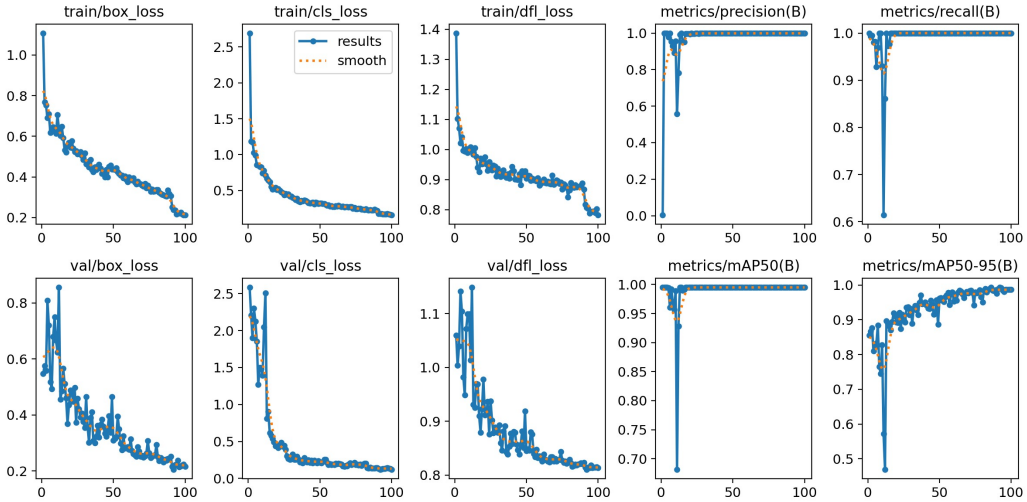

図 2: YOLOv8n モデルの紙芝居舞台検出タスクにおけるファインチューニング学習曲線．上段左から順に訓練時のバウンディングボックス損失，分類損失，DFL 損失，適合率(B)，再現率(B)，下段は検証時の各対応指標と mAP50(B)，mAP50-95(B)を示す

2.4.シーン変更点の検出

本項目は今年度卒業研究のテーマのひとつに設定し卒業研究指導として取り組むことができたことをはじめに述べておく[8].

紙芝居舞台領域の画面領域を矩形状の注目領域（Region of Interest 以降,ROI)とし当該領域内の画素オプティカルフローから水平方向移動速度を算出する．画面引抜動作の特性として，引抜開始時に ROI の全画素が同時に動き始め，徐々に次画面の静止画素が増加し，次画面完全露出時点で全画素が静止状態に移行する．この速度変化パターンを利用し，閾値ベースの変化点検出アルゴリズムを適用した．具体的には，動画フレーム内で一定期間の静止状態を検知した場合，その静止区間の開始点をシーン変更点として検出する手法を実装した．

2.5 実験結果と考察

提案手法の検出性能について分類評価と回帰評価をおこなった．分類評価では，推定結果と GrandTruth のマッチング結果から混同行列を作成した．近傍の閾値を 2 秒とし，推定値に対して GrandTruth が近傍に存在する場合を TruePositive，それ以外を FalsePositive とした．推定値以外のフレームで GrandTruth が近傍に存在しないフレーム TrueNegative，それ以外を FalseNegative とした．回帰性能評価は，TruePositive についてマッチングした GrandTruth との平均絶対誤差（MAE）と二乗平均平方根誤差（RMSE）確認した．結果を表 1 に示す．シーン変更点検出システムの評価結果は，高い検出性能と時間精度を示した．再現率と適合率はシステムが大多数の実際のシーン変更点を正確に検出し，誤検出を抑えていることを示唆する．時間的精度においては，0.28 秒という平均絶対誤差は実用的な検出精度を達成しており，紙芝居ストーリーテリング分析の文脈では許容範囲内であるが，RMSE 値が MAE より若干大きい点は，一部検出点に比較的大きな時間的誤差が存在することを示唆している．検出失敗の例は，画面の前面と背面の色彩およびテクスチャが類似している場合に生じる．画像の色彩やテクスチャに起因する誤検出は，固定された閾値や前提条件に依存するルールベース認識手法において顕著に発生する．

表１　性能評価

再現率 (Recall)	適合率 (Precision)	F 値 (F-mearsure)	平均絶対誤差 (MAE)	二乗平均平方根誤差 (RMSE)
0.98	0.93	0.95	0.28sec	0.40sec

3.紙芝居統合タイムライン可視化システム

3.1.提案手法

演者・観客の顔向きを統合可視化するシステムを提案する．提案手法の概要を図 3 に示す．システム構成は，演者・観客の映像入力から，InsightFace[9]エコシステムを用いた 3 軸（Yaw，Pitch，Roll）顔向き推定，時系列データ統合，可視化出力の各モジュールで構成される．顔向き推定処理は CUDA 対応 GPU 上でオフライン実行され，観客分析では初期フレームを用いた顔検出範囲および個体 ID 割当を事前設定する．可視化機能は，全対象 ID の顔向き統計量（基本統計値，角度別頻度分布，角度分布散布図）および個別 ID 毎の時系列角度変化グラフを生成する．これにより演者・観客間の視線行動を時系列でラベリングしたタイムラインが得られ，紙芝居上演における行動分析のための定量的データセットを提供する．

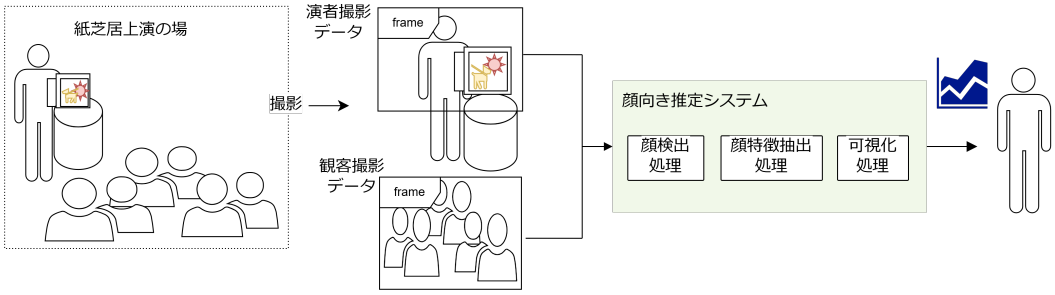

図 3. 紙芝居統合タイムライン可視化システム概要

3.2.実験

本研究では倫理的配慮の下，保育園児および高齢者を対象とした紙芝居上演の撮影を実施した．被験者の匿名性保持を条件として研究目的のみの利用許諾を得た．本報告では，マスク着用のない園児を対象とした撮影データを分析対象とした．実験環境として，「るるのおうち」（童心社制作，脚本・絵：まついのりこ）全 12 場面の紙芝居を使用し，観客構成は園児 38 名，大人 2 名（うち 1 名は途中参加）であった．顔トラッキングの初期設定では観客 39 名（園児 38 名，大人 1 名）に対し，ID 0〜38 を割り当てた．映像処理は縦解像度 640 ピクセル，サンプリングレート 1FPS で実施した．

3.3.実験結果と考察

動画フレームに検出結果情報として，顔検出結果を矩形領域，眉間から鼻頭までのベクトル，Yaw 値と Pitch 値を重畳表示した出力結果例を図 4 に示す．Yaw 値と演者の行動の二値コーディング値(1:開幕・画面引抜・閉幕，0：その他)の時系列を図 5 に示す．Yaw 値時系列データから抽出した顔向き方向（正面・左・右など）を離散的ラベルへ変換し，全取得フレームに対する各方向の出現頻度を算出した．演者および観客の各方向の占有率を表 2 と表 3 に示す．これにより観客の注視行動を時間的推移と共に定量的に評価可能となり，各方向の占有率から観客の注目傾向を客観的に分析することができる．

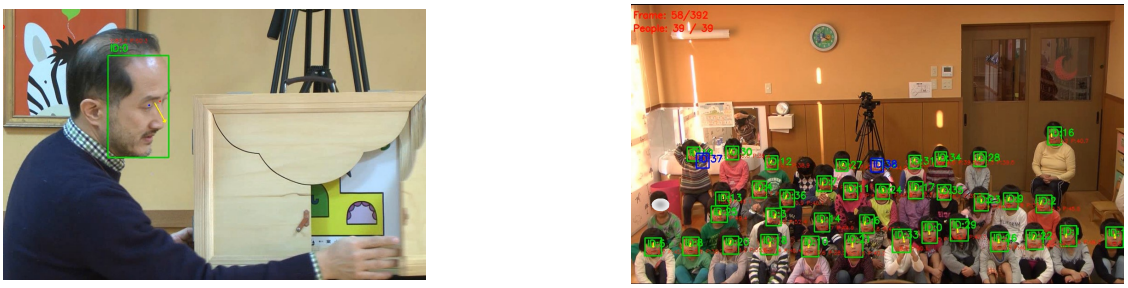

図 4.処理後の撮影データ出力結果例．左は演者，右は観客．

図 5. Yaw 値と演者の行動の二値コーディング値(1:開幕・画面引抜・閉幕，0：その他)の時系列．上段は演者，下段は観客 ID #0.

表 2. 各方向の占有率(演者)

顔向き	割合(%)
右(舞台方向)	70.7
正面(観客方向)	26.2
左(舞台と逆の方向)	3.1
欠損	0

表 3. 各方向の占有率(観客 ID 数 n=39)

顔向き	割合(%)
正面	89.1
左	22.1
右	19.0
欠損	5.9

顔向きを正面，右，左向きに分類し，図 5 と同様に Yaw 値と演者の行動の二値コーディング値(1:開幕・画面引抜・閉幕，0：その他)を演者とすべての ID に対してタイムライン形式で統合して提示することができる．図 6 では，可視化例として ID#0 から#9 に対する結果を示す．正面判定の閾値はカメラ方向を中心軸として Yaw 値±25 度とした．このタイムラインから，時間軸でスライスして参加者の行動を個別の振る舞いとして，あるいは分布という形で集団の振る舞いとして確認することを可能とする．

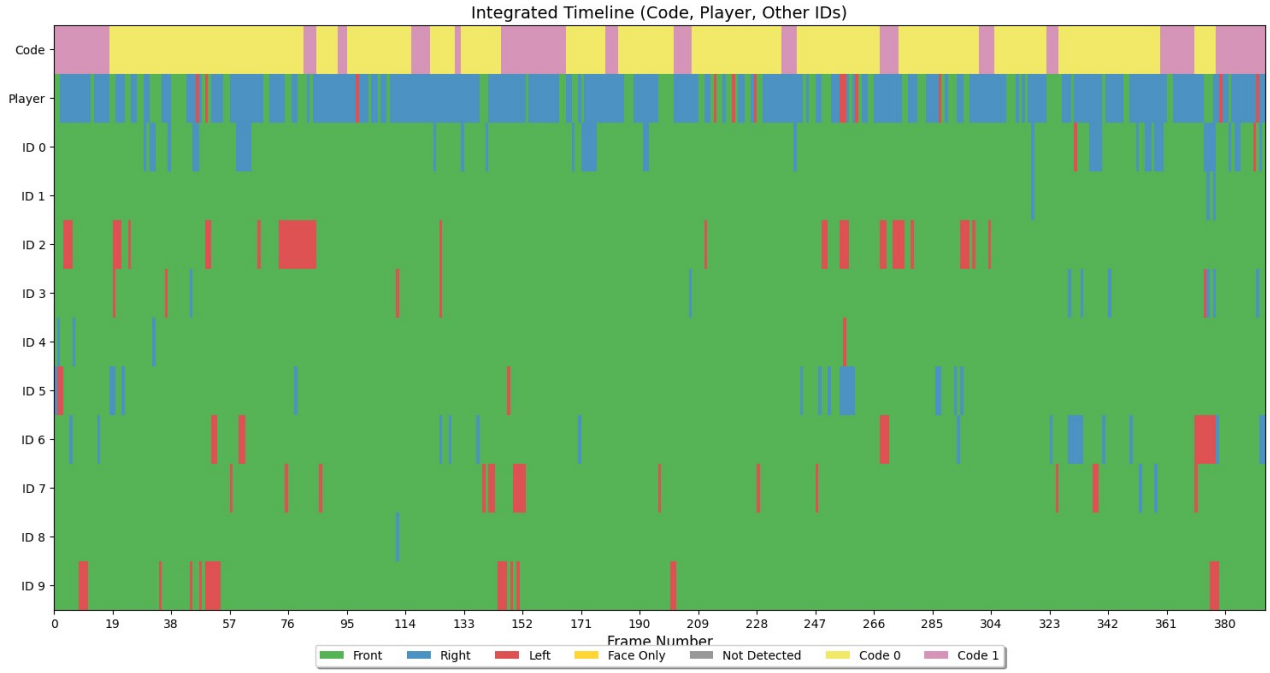

図 6.統合タイムライン結果例．演者行動コード(1:開幕・画面引抜・閉幕，0：その他)，演者顔の向き，観客顔の向き(結果例として，ID#0 から#9 を示している)

4.おわりに

紙芝居における物語進行時の演者・観客行動を自動取得し定量的に可視化することを目的として，紙芝居抜差し動作の映像データセットを作成し，紙芝居ストーリーテリングタイムラインの進行指標となる紙芝居画面が更新されるタイミングの自動検出手法を提案した．また，顔向き検出技術を用いた演者のアイコンタクト動作および観客の注目状態推定手法を提案し，実際の上演を撮影して，撮影データから，演者および観客の紙芝居ストーリーテリングタイムラインを統合した形で可視化して提示することができた．本提案にもとづく可視化手法をもちいることで，演者のアイコンタクト動作と観客の注目状況を集団と個別の両面から振る舞いを確認することができ，本手法の有効性を示唆する結果を得た．

謝辞

紙芝居上演および撮影にご協力いただいた皆様に深く感謝する．また，令和 6 年度学長裁量研究費の助成により，映像処理のための計算環境を整えることができ本研究課題に取り組むことができた．ここに深く感謝する．

参考文献

[1] 内田貴久,港隆史,石黒浩: “対話アンドロイドに対する主観的 意見の帰属と対話意欲の関係,” 人工知能学会論文誌, Vol.34, No.1, B-I62_1-8, 2019.

[2] 楊潔,菊池浩史,菊池英明: “雑談音声対話システムによる繰返し発話の多様性がユーザの共感と対話継続欲求に与える効果”,日本知能情報ファジィ学会誌,Vol.36,No.4,pp.713-721（2024）.

[3] K.Yang, H.Xu, K.Gao.,Cross-modal bert for text-audio sentiment analysis, ACM Multimedia 2020, pp. 521-528 (2020).

[4] H.Sugiyama,M.Mizukami,T.Arimoto,H.Narimatsu,Y.Chiba,H.Nakajima,Empirical

Analysis of Training Strategies of Transformer-Based Japanese Chit-Chat Systems,In 2022 IEEE Spoken Language Technology workshop(SLT),2022.

[5] Z.Xu,Y.Gao.,Research on cross-modal emotion recognition based on multi-layer semantic fusion,Mathematical Biosciences and Engineering 2024,Vol.21,Issue.2:2488-2514.doi:10.3934/mbe.2024110

[6] J.Redmon,S.Divvala,R.Girshick,A.Farhadi.,You Only Look Once:Unified, Real-Time Object Detection,IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2016.

[7] 音声資源コンソーシア, 日本語単一話者オーディオブック・紙芝居朗読音声コーパス (J-KAC)https://research.nii.ac.jp/src/J-KAC.html

[8] 野村来喜.,“オプティカルフローを用いた紙芝居シーン変更点の検出”,令和 7 年度長崎県立大学情報システム学科卒業論文

[9] J.Deng,J.Guo,N.Xue,S.Zafeiriou., ArcFace: Additive Angular Margin Loss for Deep Face Recognition.,IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2019.