卒　業　研　究　報　告　書

令和7年度

研究題目

アバターの自然さ向上を目的とした
トラッキングデータ欠損補間手法

指導教員　　上　野　秀　剛　　准教授

氏　　名　　奥　田　祥　太

令和8年2月25日　提出

奈 良 工 業 高 等 専 門 学 校　　情 報 工 学 科

アバターの自然さ向上を目的とした トラッキングデータ欠損補間手法

上野研究室　奥田祥太

近年，VTuberに代表されるアバターを介したライブ配信が普及し，iPhoneのTrueDepthカメラとARKitによるフェイストラッキングが広く利用されている．一方で，フレームアウト，遮蔽，過度な頭部回転によりトラッキングが一時的に破綻し，BlendShape係数や頭部姿勢といった時系列データが欠損する場合がある．このような欠損は，アバターの停止や復帰時の跳躍を引き起こし，視聴者が知覚する自然さ(Perceptual Naturalness)を損なう要因となる．既存の配信支援ソフトウェアに実装されている欠損補間手法は，欠損直前値の保持やニュートラル姿勢への漸近といった単純な補間により欠損区間を埋める．このとき，速度や方向などの欠損前後の運動文脈を考慮しないため，欠損境界で速度不連続や不自然な収束が生じやすい．本研究はARKitによるフェイストラッキング系列を対象とし，欠損前後の短時間コンテキストから端点速度を推定し，リアルタイムに3次エルミートスプラインを用いて欠損区間を補間する手法を提案する．提案手法は指数重み付き移動平均により差分系列を平滑化して運動傾向を推定し，欠損境界における位置および速度の連続性（C^1連続）を保証することで，復帰時の飛躍や不自然な収束を抑制する．また，復帰側のコンテキストを参照するため固定遅延バッファ（$T_{\mathrm{delay}} = 3.0\,\mathrm{s}$）を導入する．提案手法の有効性を検証するため，被験者18名による主観評価実験を実施した．実際のトラッキング破綻を収録した4シナリオ(Frame-out, Occlusion, Dynamic Occlusion, Head Rotation)，計16パターンに対し，補間なし，従来手法，提案手法の3条件を適用した計48本（各10秒）の刺激動画を提示し，頭部，表情，動き全体の自然さを5段階で評価させた．Friedman検定とBonferroni補正付きWilcoxon符号順位検定の結果，提案手法は全シナリオの平均で補間なしに対しすべての評価尺度で有意に不自然さを低減し，従来手法に対しても頭部および動き全体で有意な改善を示した．さらに，シナリオ別の分析においても頭部および動き全体では一貫した改善が確認され，復帰時の姿勢変化が大きい条件ほど効果が顕著だった．以上より，欠損補間を真値の復元ではなく知覚的自然さの維持として設計し，欠損境界における速度連続性を担保することが，VTuber配信における視聴体験の改善に有効であることを示した．

目 次

1 はじめに

ライブ動画配信は，近年，世界規模の経済的・社会的現象へと発展しており[1]，その中でも，配信者が実写ではなくアニメーション化されたアバターを介して配信を行うVTuber（Virtual YouTuber）文化は，2016年に日本で誕生して以降，国際的な人気を獲得してきた[2].

VTuber配信ではそのアバターや自身のペルソナをより魅力的かつ豊かに表現することを重要としており，様々なハードウェア，ソフトウェア，デザインツールを使用して高品質なVTuberライブストリーミングを行っている[3]. なかでも，VTuberの表現の根幹を支える技術の1つであるフェイストラッキングには，iPhoneのTrueDepthカメラを用いたARKitや，Webカメラなどの民生レベルのハードウェアやソフトウェアが良く用いられている[3]. これらの手法は，産業用モーションキャプチャシステムと異なり，単眼あるいは限られたセンサー情報からトラッキングを行うため，取得精度や頑健性に制約がある. その結果，演者が画角外に出る，遮蔽物で顔が隠れる，過度な頭部回転が生じる等の要因により，トラッキングが一時的に失敗し，時系列データが欠損しやすい. この欠損は，アバターの一時停止のみならず，トラッキング復帰時の跳躍や姿勢の急変といった時間的不連続を引き起こし得る. 人間の動作を表すアニメーションが観察者に「意図通りに妥当」と知覚されるためには，文脈整合性や内的整合性（顔と身体の整合など）が重要である[4]. 特に，高い写実性を持つアバターにおいては，データ欠損に起因するわずかな動作の不整合やジッターが「不気味の谷」現象を誘発し，アバター全体への受容性を著しく低下させる要因となり得る. また，欠損に起因する挙動の破綻が顕在化しやすい現状では，配信者は画角外への移動や頭部姿勢を常に意識するなど，トラッキング欠損の発生を過度に考慮したパフォーマンスを行わざるを得ない場合がある. このような状況は，演技やライブ性への集中を妨げ，表現の自由度を制限する要因となり得る.

視聴者が違和感なく受容できる知覚的自然さ（Perceptual Naturalness）を向上するために，LuppetX[1]などの配信支援ソフトウェアにはトラッキングデータに欠損が発生した際のアバターの動作を補間する機能が実装されている. これらの機能の多くは欠損直前の状態を保持したり，固定値の秒数をかけてニュートラル姿勢へ漸近させる方式であり，欠損直前までの運動速度・方向と，復帰直後の状態を運動文脈として整合的に接続できない場合がある. このような補間を行うと，欠損境界でアバターの運動速度・方向の変化が不連続となり，表示位置の飛躍や不自然な収束が生じやすい. VTuber配信におけるアバターの動作には，視聴者が参照可能な正解動作が存在せず，また民生環境において高精度な真値復元を行うことは現実的ではない. そのため，復元精度よりも，欠損前後を通して一貫した挙動として知覚できる連続性が重要であり，本研究では知覚的自然さ（Perceptual

[1]Luppet Technologies Inc., LuppetX. `https://luppet.jp/`. （accessed 2025-12-20）

Naturalness）の維持・向上に主眼を置き，欠損補間方法を設計する．

本研究はARKitによるフェイストラッキングを対象に，トラッキングデータの欠損を欠損前後の短時間コンテキストを利用して滑らかに補間する手法を提案し，被験者実験で有効性を検証する．提案手法は，欠損前後の短区間から運動傾向を推定し，指数重み付き移動平均によりノイズを抑えつつ変化へ追従する接線の推定を行う．そのうえで，3次エルミートスプライン補間を用いて欠損境界における位置と速度の連続性（C^1連続）を保証し，従来方式で問題となりやすい復帰時の飛躍や不自然な収束を回避する．提案手法によってVTuber配信者はトラッキングの欠損を過度に意識することなくパフォーマンスを行いやすくなるとともに，アバターの知覚的自然さを向上させることが可能となる．また，ライブ配信自体が持つ一定の遅延[5]を利用し，バッファ区間を導入することで欠損後の情報を利用しながら遅延を最小化している．

以下，2章では関連研究について述べ，3章でVTuberシステム構成，ARKitによるフェイストラッキングおよびスプライン補間の数学的基礎を説明する．4章で提案手法であるコンテキスト考慮型スプライン補間の詳細を述べ，5章で提案手法を評価するための被験者実験の設計と手順を説明する．6章では実験の結果と考察を示し，7章では本研究のまとめと今後の発展について説明する．

2 関連研究

2.1 VTuber配信の運用実態と技術的課題

VTuber配信は，アバター表現・トラッキング・配信ソフトウェア・各種制作ツール等の多要素からなる実運用システムであり，品質確保には配信者の技術的負担と運用上のリスクが伴う．Kimらは，プロVTuber 16名へのインタビューおよび使用機材・ツールの調査により，アバター制作から日々の配信運用までの実務的プロセスを整理した[3]．同研究では，高品質なパフォーマンスの提供には，魅力的なペルソナの設計に加え，複数のハードウェア・ソフトウェア・制作ツールを適切に扱う配信者の技術的スキルが重要であることが指摘されている．また，ツールの統合や自動化による負担軽減の必要性についても述べられている．また，Freemanらの研究によると，Social VR空間で活動する配信者は，視聴者と自然かつ直感的に関わるために，全身運動と表情のリアルタイムトラッキングを積極的に用いることが可能であることが示唆された[6]．さらに，Luらは，1年以上継続してVTuberを視聴する視聴者21名へのインタビュー調査により，視聴動機や実写配信者との認識差，多層的なアイデンティティの理解などを明らかにした[2]．同研究によると，視聴者はVTuberの表現を細やかに解釈し，アバターと配信者の関係性を複層的に捉えることを示している．これらの知見は，リアルタイムトラッキング品質が配信体験の中核となり得ること，民生環境の運用では一時的な破綻を含む品質劣化が避け難いこと，視聴者が表現のわずかな変化も意味づける可能性があることを示す．したがって，追加機材や大幅な運用変更を要せず，欠損等の一時的破綻に起因する品質低下を緩和する支援技術には実運用上の需要がある．本研究はこの文脈において，欠損時挙動という具体要因に焦点化し，視聴者が受容できる自然さの維持・向上を狙う低遅延補間を検討対象とする．

2.2 アニメーションにおける自然さの知覚

本研究は欠損補間の評価指標として「真値への厳密復元」ではなく,「視聴者が知覚する自然さ」を扱う．Hwangらは，人工物・デジタル環境における自然さを整理し，知覚的自然さ（Perceptual Naturalness）と概念的自然さ（Conceptual Naturalness）を区別して論じている[7]．本研究が対象とする欠損時挙動の不連続は，視覚的な処理流暢性（Processing Fluency）に影響するため，主として知覚的自然さの問題として扱うのが妥当である．また，Etemadらは，人間動作アニメーションが観客に「意図通りに妥当」と知覚されるための要素を体系化し，知覚的妥当性（Perceptual Validity）という枠組みを提案している[4]．同枠組みが含む文脈依存性や内的整合性の観点は，欠損補間により顔パラメータのみが不自然に変化すると，顔と身体の内的整合性が崩れ，違和感が増幅する可能性を示唆する．評価方法の観点でも，数学的指標のみでは補間手法間差を十分に説明できない可能性がある．補

間や合成を扱う研究では，人間による知覚評価の必要性が繰り返し指摘されており[8, 9]，Leiらは，キャラクター上半身アニメーションの知覚的自然さが，複数の補間手法間で変わりうることを示した[8]．さらに，観察者の不自然さ検知はキャラクターの外見や生物らしさに依存することが報告されており[10, 11]，キャラクターの擬人性（animacy）が高い表現ほど運動は単に敏感になりうる．VTuberアバターは一般に擬人化度が高い表現を採用することが多く，欠損境界の不連続が検知されやすい可能性があるため，本研究で「視聴者が感じる自然さ」を主要評価軸に置く理由となる．加えて，Welbergenらの研究によって，バーチャルヒューマンのリアルタイムアニメーションでは，自然さと制御性（および計算量）がトレードオフとなり得ることが整理されており[12]，配信という実運用制約下で低遅延・軽量に自然さを改善するという本研究の設計方針とも整合する．

2.3 アバター支援とトラッキング品質改善

リアルタイムアバター表現の品質改善に向けては，表現支援，トラッキング品質向上，欠損補間・復元といった観点から多様な研究がある．表現支援では，Tangらが，配信者の動きを忠実に再現する従来のミラーリングではなく，配信者とアバターの結合を「緩める」ことで，キャラクターらしい表現を自動生成するシステムAlterEchoを提案した[13]．同研究では，エネルギーレベルと外向性というパラメータにより性格演出を調整し，既存ソフトウェア（VMagicMirror）との比較調査において，魅力的・自然・好ましいといった指標で高い評価を得たことが報告されている．また，トラッキング品質向上の分野では，Chenらが深層学習を用いて微細な表情変化を高忠実度に追跡するリアルタイム手法を提案し，従来よりも豊かな感情表現を捉えることを目指した[14]．これらは表現品質を向上させる一方で，遮蔽やフレームアウト等に起因する欠損が完全に解消されるとは限らず，破綻が生じた瞬間の品質低下を抑える補助的手段も依然として必要となる．したがって，本研究は，表現の生成・誇張や追跡の高忠実度化そのものではなく，トラッキングが破綻した瞬間の品質低下を抑える補助的技術として，これらの技術と併用可能である．

欠損補間・復元については，モーションキャプチャや運動計測における欠損区間の補間が重要な課題である．Bradwellらはモーションキャプチャ駆動アニメーションの解説において，遮蔽による欠損やジッターが起こり得ることを指摘し，欠損は補間やスプラインで埋め，ノイズはフィルタで平滑化する必要があることを述べている[15]．Howarthらは，人間動作解析における欠損補間手法を比較し，欠損時間が短い場合には3次スプラインが有効である一方，欠損が長い場合には別手法が有利となるなど，欠損長に応じて適切な補間が変化することを示した[16]．ただし，これらの研究は主に計測精度の観点から評価されており，視聴者の知覚やリアルタイム制約は前提としない．逐次入力されるストリーミングデータに対

する補間では，全データを前提とするグローバルなスプラインは扱いにくく，局所情報で導関数を推定してC^1連続な曲線を生成する設計が重要となる．Debskiらは，逐次入力されるデータに対してリアルタイムに3次スプライン補間を行う手法を提案し，特にエルミートスプラインを用いる場合には，一次導関数（傾き）推定が性能を左右することを指摘している[17].

以上を踏まえると，表現支援や追跡高忠実度化，補間理論・実装といった先行研究は一定数存在する．一方で，民生環境のVTuber配信における欠損シナリオを想定した補間手法や，さらにその効果を視聴者の知覚的自然さで体系的に評価した例は多くない．本研究は，欠損前後の短時間コンテキストから端点速度を推定し，3次エルミートスプラインにより欠損境界でのC^1連続性を保証することで，欠損中・復帰時の不連続を低遅延に緩和する．さらに，欠損発生シナリオの違いを考慮した主観評価実験により，自然さの改善がどの条件で一貫して観測され，どの条件で変動し得るかを明らかにする点に新規性がある．

3 準備

3.1 VTuberシステムの技術構成

本研究が対象とするフェイストラッキング主体のVTuber配信は以下の3段階を経て作成される.

1. トラッキング
   iPhoneやWebカメラ等のデバイスにより顔の形状・表情・頭部姿勢を推定し, BlendShape係数や頭部姿勢(位置・回転), 視線方向等のパラメータ列として取得する.

2. アバター制御・描画
   パラメータをUnity等のレンダリング環境で2Dまたは3Dアバターへ適用し, 配信者の表情・姿勢に追従するアバターアニメーションを生成する.

3. 映像出力(配信)
   生成した映像をOBS Studio [2]に代表される配信ソフトウェアへ入力し, 視聴者に対してライブ映像として提示する.

本研究は1.で取得されたパラメータが2.でアバター制御へ入力される直前に着目する. 演者の姿勢変化や遮蔽等によりトラッキングデータが欠損するとアバターアニメーションが生成されず, アバターの動作が停止したまま映像が出力される. 提案手法はアバター制御・描画段階に渡すパラメータ値の欠損区間を補間することで欠損によるアバターの動作の停止を抑制し, 知覚的自然さを改善する.

配信の実運用では, トラッキングデータをUnityなどの描画環境へ直接入力する構成に加えて, 配信支援ソフトウェアを用いる構成も広く採用されている. 配信支援ソフトウェア(例:LuppetX, VMagicMirror[3], Animaze[4]など)は, 単一のアプリケーション内でフェイストラッキング(または外部からの受信), トラッキング結果に基づいたアバターのレンダリング, 配信ソフトウェアへの出力を一括して担う. 配信支援ソフトウェアの機能の1つとして, トラッキングデータの平滑化・補正や, トラッキング欠損時の挙動制御が提供される場合がある. 本研究では, トラッキングデータ補間の従来手法として配信支援ソフトウェアであるLuppetXの実装を参考とした, 典型的な欠損時挙動制御を模した方式を採用する. 具体的なアルゴリズムについては, 3.4節で述べる.

[2]OBS Project, OBS Studio. https://obsproject.com/ja/. (accessed 2026-01-18)
[3]獏星 (ばくすたー), VMagicMirror. https://malaybaku.github.io/VMagicMirror/. (accessed 2026-01-18)
[4]Holotech Studios, Inc., Animaze by Facerig. https://www.animaze.us/. (accessed 2026-01-18)

表 1: フェイストラッキングデータの概要

データ種別	内容
表情 (BlendShape 係数)	口の開閉, まばたき, 眉の上下, 口角の引き上げ, 頬の膨らみ等の顔表情の変化を表す52種類のパラメータ群.
頭部姿勢	頭部の三次元的な位置および回転を表すパラメータ
視線	左右眼球の回転方向を表すパラメータ

3.2 フェイストラッキング

フェイストラッキングは, カメラで撮影した人間の顔の形状や動きから表情や頭部運動をパラメータとして取得する技術であり, VTuber のアバターアニメーションの根幹となる技術である. 本研究では, フェイストラッキングの具体的な実装例として Apple が提供する AR フレームワークである ARKit を用いる. ARKit は TrueDepth カメラ等を用いて顔の形状および動作をリアルタイムに推定し, アプリケーションへ出力する機能を備えている. 表1に本研究で扱う, ARKit が出力するフェイストラッキングデータを示す. これらの値はフレームごとに逐次取得される時系列データであり, 遮蔽や姿勢変化等により一部フレームで値が得られない場合を欠損として取り扱う.

欠損区間ではフェイストラッキング値が得られないため, 実装上は欠損直前値の保持や3.4節に示す従来手法によるニュートラルへの収束など, 代替値でアバターをアニメーションさせる. この間に演者が動くと, 演者の状態とアバターの位置や回転などのずれが蓄積し, 欠損終了時に観測されたトラッキングデータへ一括で更新されることで, 姿勢や表情の急変が生じ得る. 一方, 欠損中に演者がほとんど動かなければ, ずれが小さく跳躍は顕在化しない. また, 欠損中の停止や欠損終点での急加速により, 速度の不連続として不自然さが知覚される場合がある. 頭部姿勢や視線などの回転量も同様に, 欠損終点での急激な角度変化が不自然さとして現れ得る.

3.3 3次エルミートスプライン

欠損区間を滑らかに埋めるため, 本研究では3次エルミートスプラインを用いる. 3次エルミートスプライン (Hermite Spline) は, 2点間を結ぶ曲線を, 端点の位置に加えて端点における接線 (1次微分) を指定して定める補間法である. 曲線補間の文脈では, 1次微分は端点での運動の向きや変化傾向を表現する条件として用いられる. そのため, 3次エルミートスプラインでは端点位置と接線条件を与えることで, 欠損境界で観測系列と接続する際に位置および1次微分が連続となる C^1 連続性を満たすな曲線を構成できる. この性質から, 滑らかな運動軌跡の生成やキーフレーム補間など, 平面・3次元空間上の曲線生成に広く用いられる. 本研究においても, 欠損境界における見た目の不連続を抑制することを目的として, 3次エルミートスプラインを欠損補間に用いる.

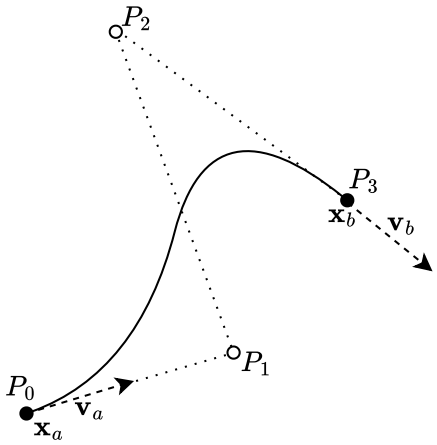

図 1: 3次エルミートスプライン

図1に3次エルミートスプラインの例を示す. 黒実線は補間によって生成される曲線である. $\mathbf{P}_0$ および $\mathbf{P}_3$ は補間区間の始点および終点位置 $\mathbf{x}_a, \mathbf{x}_b$ に対応する. $\mathbf{v}_a, \mathbf{v}_b$ は, それぞれ始点および終点における曲線の接線（1次微分）を表し, 図中では端点から伸びる矢印として示されている. エルミートスプラインでは, これらの端点における接線条件を与えることで, 曲線が端点でどの方向に接続されるかが決定される. 図中の点線は, これらの端点条件を3次ベジエ曲線として表現した際の制御点 $\mathbf{P}_1, \mathbf{P}_2$ への対応関係を示している.

以下では, 各時刻のトラッキングデータを d 次元ベクトル $\mathbf{x}(t) \in R^d$ として扱う. 区間 $[t_a, t_b]$ ($t_a < t_b$) において, 端点の位置を $\mathbf{x}_a = \mathbf{x}(t_a)$, $\mathbf{x}_b = \mathbf{x}(t_b)$, 端点の接線を $\mathbf{v}_a = \frac{d\mathbf{x}}{dt}\big|_{t=t_a}$, $\mathbf{v}_b = \frac{d\mathbf{x}}{dt}\big|_{t=t_b}$ とする. 区間長を $T = t_b - t_a$ とおき, 正規化時間 $s = (t - t_a)/T \in [0, 1]$ を導入すると, 3次エルミート補間は式 (3.3.1) で与えられる.

$$\mathbf{x}(s) = h_{00}(s)\mathbf{x}_a + h_{10}(s)\,T\mathbf{v}_a + h_{01}(s)\mathbf{x}_b + h_{11}(s)\,T\mathbf{v}_b, \tag{3.3.1}$$

ただし

$$h_{00}(s) = 2s^3 - 3s^2 + 1, \quad h_{10}(s) = s^3 - 2s^2 + s, \quad h_{01}(s) = -2s^3 + 3s^2, \quad h_{11}(s) = s^3 - s^2 \tag{3.3.2}$$

である. この表現により, $\mathbf{x}(0) = \mathbf{x}_a$, $\mathbf{x}(1) = \mathbf{x}_b$, および $\frac{d\mathbf{x}}{dt}\big|_{t=t_a} = \mathbf{v}_a$, $\frac{d\mathbf{x}}{dt}\big|_{t=t_b} = \mathbf{v}_b$ が保証される.

また, 同じ曲線は3次ベジエ曲線としても表現できる. 3次ベジエ曲線 $\mathbf{B}(s)$ を制御点 $\mathbf{P}_0, \mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$ により式 (3.3.3) と定義する.

$$\mathbf{B}(s) = (1-s)^3\mathbf{P}_0 + 3(1-s)^2 s\mathbf{P}_1 + 3(1-s)s^2\mathbf{P}_2 + s^3\mathbf{P}_3, \quad s \in [0, 1] \tag{3.3.3}$$

エルミート形式の端点条件を満たす制御点は式 (3.3.4) で与えられる.

$$\mathbf{P}_0 = \mathbf{x}_a, \qquad \mathbf{P}_3 = \mathbf{x}_b, \qquad \mathbf{P}_1 = \mathbf{x}_a + \frac{T}{3}\mathbf{v}_a, \qquad \mathbf{P}_2 = \mathbf{x}_b - \frac{T}{3}\mathbf{v}_b \tag{3.3.4}$$

本研究では, 3次エルミート補間を用いて欠損前後の短時間コンテキストから端点の接線条件 $\mathbf{v}_a, \mathbf{v}_b$ を推定し, 欠損区間を補間する.

3.4 従来手法

本研究では，提案手法の有効性を検証するための比較対象として，配信支援ソフトウェアの1つであるLuppetXにおけるトラッキング欠損時の挙動を参考にする．特定ソフトウェア固有の実装詳細に依存した議論を避けるため，同ソフトウェアの挙動を厳密に再現するのではなく,「欠損中は値を一定期間保持した後，ニュートラルへ収束させ，復帰時に観測値へ戻す」という振る舞いをアルゴリズムとして抽象化する．このアルゴリズムは，欠損区間の先頭で得られる欠損直前値と欠損終了後に得られる復帰値を用い，欠損区間内の各フレームの出力を次の3フェーズで生成する．

従来手法

- **フェーズ1：保持**
  欠損開始後，T_{hold} 秒間は欠損直前値を保持する（停止）．
- **フェーズ2：ニュートラルへの収束**（フェーズ1終了後〜フェーズ3開始まで）
  欠損直前値から数値的な原点 $\mathbf{0}$ へ向けて，係数 α による指数減衰により滑らかに収束させる．ここで $\mathbf{0}$ はデータの表現空間におけるゼロベクトル（スカラー値の場合は0）を表し，本手法では一律にこのゼロ値を**ニュートラル状態**として扱う．
- **フェーズ3：復帰値への遷移**
  欠損区間の終端 T_{trans} 秒では，フェーズ2の最終値から復帰値へフェーズ2と同様に指数減衰により滑らかに収束させる．

上記の従来手法の各種パラメータはLuppetXの実装を参考に，$T_{\mathrm{hold}} = 0.3\,\mathrm{s}$, $T_{\mathrm{trans}} = 0.2\,\mathrm{s}$, $\alpha = 0.4$ とした．この手法は欠損中の破綻を隠蔽しやすい一方で，欠損前後の速度やその方向などの運動文脈を利用しないため，復帰時に速度不連続や跳躍が生じ得る．

4 提案手法

4.1 概要

VTuber配信におけるフェイストラッキングデータを対象として，トラッキングデータ欠損中のアバター挙動を補間生成するリアルタイム手法を提案する．提案手法が補間対象とするトラッキングデータの成分は，以下の5種類である．

- BlendShape係数（52種類）
- 頭部位置
- 頭部向き
- 左目眼球回転
- 右目眼球回転

提案手法の処理概要を図2に示す. 矢印が処理とデータの流れを示しており，青色で示した領域が本研究で提案する欠損補間処理の適用範囲である．入力として，iPhoneによるARKitフェイストラッキングから，頭部位置，頭部回転，視線方向，およびBlendShape係数が逐次取得される．提案手法では欠損復帰後の観測値を補間に利用するため，入力パラメータをバッファに蓄積して固定遅延T_{delay}を導入し，遅延参照後の系列に対して欠損検出および補間処理を行う. トラッキングデータの欠損が検出された場合，欠損直前および欠損復帰直後の短時間コンテキストから端点速度を推定し，3次エルミートスプラインにより欠損区間の軌道を生成する．なお，本研究では欠損の検出方法は提案せず，任意の方法でトラッキングデータの欠損区間が与えられることを前提とする．補間後のパラメータは通常のレンダリング処理に入力され，最終的にライブ配信される．

補間アルゴリズムとして3次エルミートスプラインを採用した理由を述べる．欠損区間の補間には，端点の位置と速度を境界条件として指定しC_1連続性を保証できること，局所情報のみでリアルタイムに曲線を構成できること，計算コストが比較的低いことの3つを要件とした．通常の3次スプラインは全データ点を前提とするため逐次処理に適さない．Catmull-Romスプラインは接線が隣接点から一意に決まるため，柔軟な接線設計が難しい．線形補間はC_1連続性を保証できない．3次エルミートスプラインはこれらの要件をいずれも満たすため，本研究の補間アルゴリズムとして採用した．

本実装では$T_{\mathrm{delay}} = 3.0\,\mathrm{s}$とした．代表的な動画配信サイトYouTubeにおいては通常2sから15s程度の遅延が発生している．本手法が導入する追加の3.0sの遅延は，このプラットフォームが標準的に抱える遅延の範疇に収まるものであり，リアルタイムの応答性への影響は限定的である．

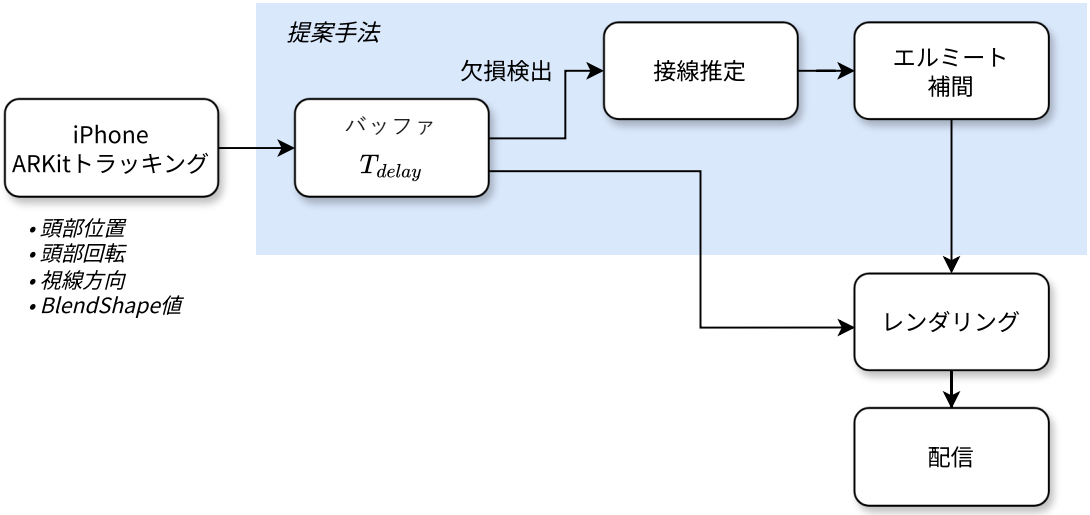

図 2: 提案手法の概念図

この手法により，欠損区間の両端において位置と速度が連続となる（C^1 連続）ように接続でき，欠損境界で生じる視覚的な不連続を抑制する．

4.2 バッファリングと欠損検出

図3にリアルタイムに取得されるトラッキンデータの入力系列と，固定遅延T_{delay}を導入したあとにレンダリングされる出力系列の関係を示す. 上段はARKitにより逐次取得されるトラッキングデータの入力系列，下段は遅延後に実際にレンダリングされる出力系列を示す. 提案手法では欠損復帰後のトラッキングデータを補間に利用するため，入力系列をバッファに蓄積し，時刻tにおいてはT_{delay}だけ過去のフレームを参照して処理を行う.

フレーム時刻をt, フレーム間隔をΔtとし，フレームtで取得されるトラッキングデータをx_tと表す. x_tは頭部位置の3次元ベクトル成分に加え，BlendShape係数のようなスカラー成分および頭部向きや視線ベクトルのようにEuler角で表したベクトル成分を含む. 以降は，成分ごとに同一の補間手順を適用するとみなし，ベクトル値は`Vector3`列，スカラー値は同一式をスカラーへ適用することで処理する. 欠損復帰後の観測値を参照して補間を行うため，フェイストラッキングデータの取得からレンダリングまで，固定遅延T_{delay}を導入する. 実装上は入力系列をリングバッファに蓄積し，レンダリングの際は$D = \lfloor T_{\mathrm{delay}}/\Delta t \rfloor$フレーム過去の値を参照する. 以降の説明では，遅延参照後にレンダリングされる系列を処理対象とみなし，簡単のため同じ記号x_tで表す.

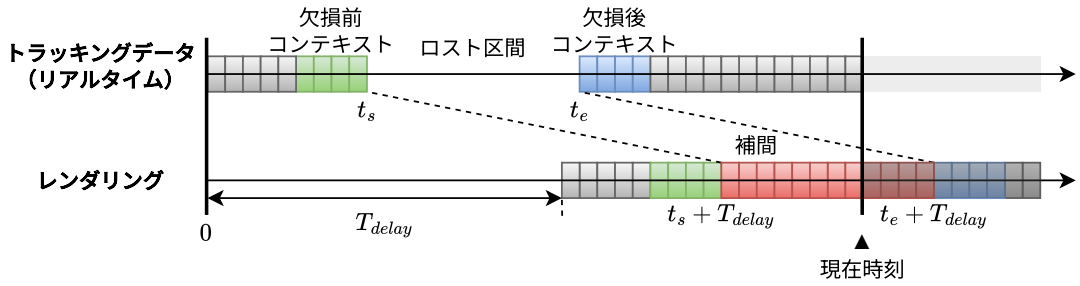

図 3: 入力系列と出力系列の時間軸

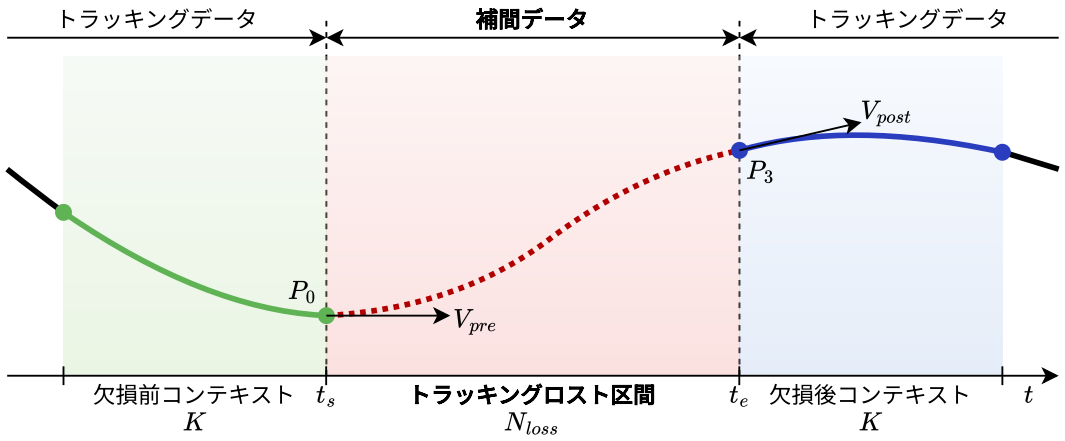

図 4: 欠損区間に対する 3 次エルミートスプライン補間の概要

4.3 接線推定

図 4 に欠損区間に対する 3 次エルミートスプライン補間の概要を示す. トラッキングデータが一時的に欠損した区間をロスト区間とし, その直前および直後のそれぞれ K フレームを欠損前コンテキスト, 欠損後コンテキストとして定義する. 欠損直前の最終フレーム p_{-1} を始点 P_0, 欠損復帰後の最初のフレーム p_0 を終点 P_3 とする. 本実装で用いた K の具体値は 5 章で述べる.

端点の接線推定には物理的な速度（値/秒）ではなく, 1 フレーム当たりの変位（値/フレーム）を用いる. 具体的には, フレーム差分 $v_i = p_i - p_{i-1}$ を計算し, これを指数重み付き移動平均で平滑化して代表差分 V_i を得る. なお, 本実装で用いた α の具体値は 5 章で述べる.

$$V_i = \alpha v_i + (1 - \alpha) V_{i-1} \tag{4.3.1}$$

欠損前後の短時間コンテキストから, フレーム差分に基づく代表ベクトル $V_{\mathrm{pre}}, V_{\mathrm{post}}$ をそれぞれ推定する. これらは, 欠損開始側および終了側における運動の傾向を表す. 欠損長（フレーム数）を N_{loss} とし, フレーム差分速度を欠損長に比例してスケーリングすることで, 欠損区間長を反映した端点接線を定義する.

$$\tilde{V}_{\mathrm{pre}} = V_{\mathrm{pre}}\, N_{\mathrm{loss}}, \qquad \tilde{V}_{\mathrm{post}} = V_{\mathrm{post}}\, N_{\mathrm{loss}} \tag{4.3.2}$$

4.4 補間

式(4.3.2)の$\tilde{V}_{\mathrm{pre}}, \tilde{V}_{\mathrm{post}}$は，正規化時間に対する端点接線として用いる．これらは，3次エルミート補間を3次ベジエ曲線として表現した際の制御点距離を決定する．

$$\mathbf{P}_1 = \mathbf{P}_0 + \frac{1}{3}\tilde{V}_{\mathrm{pre}}, \qquad \mathbf{P}_2 = \mathbf{P}_3 - \frac{1}{3}\tilde{V}_{\mathrm{post}} \tag{4.4.1}$$

このとき，補間区間長で正規化することで端点におけるフレーム当たりの速度を$V_{\mathrm{pre}}, V_{\mathrm{post}}$に一致させ，欠損前後の運動傾向との連続性を維持する．このスケーリングにより，欠損長が短い場合に過度な曲率が生じる問題や，欠損長が長い場合に変化量が不足して不自然に収束する問題を抑制し，欠損区間の長さに応じた自然な補間軌跡を生成できる．

本手法は欠損復帰後のコンテキストを用いるため，固定遅延T_{delay}の範囲を超える長時間欠損では，出力時点で復帰側速度V_{post}を参照できない区間が生じる．そこで本実装では，欠損長$N_{\mathrm{loss}} > D - K$の場合，欠損の先頭側は補間なし（その場で停止）と同様に表現し，復帰直前の$D-K$フレーム区間に対してのみ提案手法を適用する．これは視覚的に顕著となりやすい復帰時の飛躍の抑制を優先するための設計である．

5 実験

5.1 概要

提案手法が視聴者の知覚する自然さに与える影響を検証するため，被験者実験を実施する．被験者は著者らの所属する奈良高専情報工学科に在席する学生18名（16〜20歳）である．被験者は欠損を含むアバター動画を視聴し，動きの自然さを主観評価する．原因の異なる4種の欠損シナリオに基づいて作成した16種類の動画に対して，3種類の補間手法（補間なし，従来手法，提案手法）を適用した48動画を用意する．

提案手法の評価を行うために3つの研究課題（RQ）を設定する．

RQ1. 提案手法は補間を行わない場合と比較して，視聴者が感じるアバター挙動の自然さを向上させるか？

RQ2. 提案手法は従来手法と比較して，視聴者が感じるアバター挙動の自然さを向上させるか？

RQ3. 欠損発生シナリオの違いによって，各補間手法に対する自然さの評価傾向は変化するか？

5.2 評価シナリオ

本実験では，欠損区間に対する3種類の補間手法を比較する．

- 補間なし（None）：欠損区間では欠損直前の値を保持し，欠損復帰時に入力値へ即時に遷移する．
- 従来手法（LuppetX）：既存手法（3.4節）．
- **提案手法（Spline）**：欠損前後の短時間コンテキストから端点接線を推定し，3次エルミートスプラインにより欠損区間を生成する．

VTuber配信においては，フェイストラッキングが一時的に失敗し，時系列データが欠損する状況が生じ得る．本研究では，そのような欠損状況のうち，視聴者に知覚的な不連続が生じやすいと考えられる以下の4種類のシナリオを対象として評価を行う．

- Frame-out：演者の顔が画角から外れることで発生する欠損
- Occlusion：手などの遮蔽物が移動することで演者の顔が隠れて発生する欠損
- Dynamic Occlusion：演者が遮蔽物の後ろを移動することで顔が隠れて発生する欠損

- Head Rotation：頭部回転により顔の追跡が破綻することで発生する欠損

なお，Occlusion と Dynamic Occlusion はいずれも遮蔽物による欠損を扱うが，欠損発生時におけるアバターの運動状態が異なる．Occlusion では，演者自身はほぼ静止した状態で遮蔽物のみが移動するのに対し，Dynamic Occlusion では，演者が移動することで遮蔽物の背後に入り，欠損前後で頭部位置や姿勢が変化する．この違いにより，欠損前後の運動文脈の連続性や，補間に必要な端点接線の性質が異なると考えられるため，本研究では両者を別個の評価シナリオとして扱う．

各シナリオにつき 4 種類の欠損パターンを設定し，合計 16 種類の欠損パターンを用意する．表 2 に欠損パターンの一覧を示す．

5.3 刺激動画

5.3.1 トラッキング

本研究で用いる刺激動画は，欠損注入による合成ではなく，実際のトラッキング破綻により生じたデータ欠損を収録して作成する．まず，実際の演者（著者）が表 2 に示した全欠損パターンの動作を実演したものを 10 秒間でトラッキングして収録し，4.1 節に示した 5 成分のパラメータ列を記録する．欠損の発生時刻はいずれも動画開始後 1 秒から 9 秒の範囲に収まるよう設定し，各刺激動画には欠損が 2 回発生するように動作を設計する．

Frame-out では，顔がカメラ視野から外れるように上下方向および右下，左下方向へ移動する動作を行う．Occlusion では，手・腕などの身体動作により顔を上下左右から一時的に遮蔽し，追跡が外れる状況を生じさせる．Dynamic Occlusion では，画面中央の障害物に対して上下左右方向から移動しつつ遮蔽状態を作ることで欠損を生じさせる．Head Rotation では，おおよそ 90 度程度まで上下左右に回旋し，横顔相当の姿勢で追跡が破綻する状況を収録する．なお，Occlusion および Dynamic Occlusion では，実際の障害物は動画に表示されない．

表 2: 欠損シナリオと欠損パターン

シナリオ	パターン ID	内容
Frame-out	F1–F4	左／右／左下／右下方向へのフレームアウト
Occlusion	O1–O4	左／右／下／上から遮蔽物が侵入し，画面中央の顔を遮蔽
Dynamic Occlusion	D1–D4	画面中央の遮蔽物に対して，右から左／左から右／下から上／上から下に演者が移動して通過
Head Rotation	R1–R4	過度に左／右／下／上を向く頭部回転により追跡が破綻

収録中は大きな表情変化（例：口の開閉，誇張した笑顔）や視線移動を意図的に行わず，演者はリラックスした表情を維持する．ただし，瞬き等の非随意的な微小表情変化はトラッキングする．フェイストラッキングにはiPhone 15のTrueDepthカメラ用いる．また，補間とアバターのレンダリング，動画の書き出しにはUnity 6000.0.23fを用いる．

5.3.2 補間

同一の収録データに対して，3手法（補間なし・従来手法・提案手法）を適用する．この手順により，欠損の発生時刻・欠損長・元の動きが手法間で一致し，手法間比較の公平性を確保する．従来手法（LuppetX）は，欠損開始後の保持時間を0.3 s，復帰時の遷移時間を0.2 sとし，保持後はニュートラルへ指数減衰により収束させた（減衰率0.4）．提案手法（Spline）は，固定遅延$T_{\mathrm{delay}} = 3.0\,\mathrm{s}$を導入し，端点接線推定に用いるコンテキスト長は$K = 10$，指数重み付き移動平均の係数は$\alpha = 0.3$とする．本実装においては，欠損を「全成分が完全一致する入力値が$L = 10$フレーム以上連続する状態」として検出した．すなわち，$x_t = x_{t-1}$が連続して成立した回数をカウントし，Lに達した時点で欠損状態へ遷移する．

収録した全48本の刺激動画において，欠損長は平均1.04秒，標準偏差は0.39秒，最小値は0.542秒，最大値は1.909秒である．これは，欠損長$N_{\mathrm{loss}} \leq D - K$を満たし，全試行で復帰側速度$V_{post}$を参照可能であり，提案手法の端点接線推定が成立する設定となっている．

5.3.3 動画出力

各刺激動画は長さ10秒，無音，解像度は1920×1080，フレームレートは60 fpsである．背景は白の単色とし，カメラはバストアップ構図で固定した．テキストや図形などの要素は付与しない．

動画中央に表示されるアバターとしてVRoid AvatarSample A [5]を用いる．本アバターに調整を加えることでARKitのBlendShape係数が動作可能な状態とする．図5に刺激動画の代表フレーム例を示す．また，図6にFrame-out（F1）における各補間手法の欠損前後のフレーム比較を示す．上段から補間なし，従来手法，提案手法であり，左から右へ時間経過に沿って等間隔にフレームを抽出している．補間なしでは欠損中にアバターが停止し復帰時に跳躍が生じるのに対し，提案手法では欠損前後の運動方向に沿った滑らかな遷移が確認できる．

[5] VRoid Project. “AvatarSample_A .” VRoid Hub, https://hub.vroid.com/characters/2843975675147313744/models/5644550979324015604.

図 5: 刺激動画の例

図 6: Frame-out（F1）における各補間手法の欠損前後のフレーム比較

5.4 主観評価

被験者は各動画を視聴した後，以下の3項目について5段階リッカート尺度で評価する．

- 頭部：体・首周りの動きの自然さ
- 表情：表情変化の自然さ
- 全体：動き全体の自然さ

尺度は1を「とても自然」，5を「とても不自然」とし，値が大きいほど不自然さが高いことを表す．なお，実験環境上，動画再生直後や終了直前に動画切替時の描画負荷等に起因すると考えられる一時的なカクつきが一部の再生環境において観測された．これらはアバター挙動そのものに起因する現象ではないため，被験者にはこのような一時的なカクつきを評価対象外とし，それ以外の区間におけるアバター挙動の自然さを評価するよう教示する．

各項目について3種類の手法間の差の検定にFriedman検定を用いる．Friedman検定で有意差が認められた場合，事後比較としてWilcoxon符号順位検定を行い，3

表 3: 事前アンケートの質問項目

質問	選択肢
VTuberの動画・配信において，アバターの不自然な挙動が気になった経験があるか	気になったことがある／気になったことがない／わからない（VTuber動画未視聴を含む）
VTuber視聴時に，アバターの動き・表情をどの程度意識するか	非常に意識する／少し意識する／あまり意識しない／全く意識しない／VTuberの動画を見ない
不自然だと感じた動画の傾向や，気になった点	自由記述

組の多重比較に対してBonferroni補正（有意水準 $\alpha/3$）を適用する．各指標に対する評価は (i) 全刺激に対する評価，および，(ii) 4シナリオそれぞれに対する評価の2種類を実施する．RQ1とRQ2は全刺激，および，シナリオ別の結果から回答する．RQ3はシナリオ別の結果を補間手法ごとに比較することで回答する．

また，被験者のVTuber視聴経験や，アバター挙動に対する関心度が主観評価結果に与える影響を検討するため，すべての刺激動画の視聴後にMicrosoft Formsを用いた事後アンケートを実施する．質問項目を表3に示す.

5.5 実験手順

実験はPsychoPyで実装した実験プログラムにより実施する．被験者は教示文を読んだ後，練習試行で評価方法を確認し，本試行へ移行する．本試行では48本の刺激動画を提示し，各動画の視聴後に3項目を回答させる．提示順序は被験者ごとに無作為化し，順序効果および学習効果を低減した．刺激提示はすべての被験者で同一のディスプレイで行った．ディスプレイの解像度は1920×1080，大きさは24インチ，リフレッシュレートは60 Hzである．

6 結果と考察

6.1 補間による自然さの変化

表4に補間手法ごとの各動画に対する主観評価の平均値と標準偏差を示す. 補間なしと提案手法を比較すると頭部, 表情, 全体のいずれにおいても提案手法の方が平均値が低く, より自然であると評価された. 同様に, 従来手法と提案手法を比較しても, すべての評価項目で提案手法の方が平均値が低いが, 補間無しと比べるとその差はわずかに小さくなった. 表5に手法間の差に対する検定の結果を示す. Friedman検定の結果, いずれの手法においても有意差 ($p < 0.001$) が認められ, 事後比較では提案手法は補正無しに対してすべての評価項目で有意に不自然さを低減した. また, 提案手法は従来手法に対して頭部および全体で有意に不自然さを低減した一方, 表情では有意差は確認されなかった. 補正無しと従来手法では全体のみ有意差が認められた. これらの傾向は, 図7においても確認でき, 提案手法がすべての評価項目で一貫して低い評点を示したこと, および主に補間無し（ならびに頭部・全体では従来手法）との有意差が確認されたことが分かる.

全体結果として提案手法は頭部・全体で一貫して改善が観測されたため, 以下ではその差を各手法の設計上の違いから考察する. 補間なしは欠損区間で値を保持して停止し, 復帰時に入力値へ即時遷移するため, 欠損前後の境界で位置・速度の不連続が顕在化しやすい. 従来手法は欠損中にトラッキングデータをニュートラル状態へ収束させるため, アバターの一時停止や復帰時の跳躍をある程度抑制しうる一方, 欠損前の運動方向や速度を保持しないため, 欠損中の停止や収束自体が違和感として知覚される場合があると考えられる. これに対し提案手法は欠損長が固定遅延 T_{delay} を超えない限り, 欠損前後のコンテキストと C^1 連続性を満たす軌道を生成するため, 復帰時の跳躍を抑制できる. この設計差が, 頭部および全体における一貫した改善として観測されたと解釈できる. ただし, 端点速度は欠損前後の観測に基づく推定値であり, 急な表情変化や急旋回など, 欠損中に運動が急変する場合や, 欠損前コンテキストや欠損後コンテキストにそのような急変が含まれる場合, 得られる補間が真の運動文脈と大きく乖離し,

表 4: 補間手法ごとの主観評価（平均・標準偏差）

手法	頭部		表情		全体	
	平均	S.D.	平均	S.D.	平均	S.D.
補間なし	3.47	1.21	2.64	0.78	4.03	0.78
従来手法	3.22	1.14	2.36	0.61	3.53	0.79
提案手法	1.92	0.73	1.94	0.57	1.89	0.61

※ 1=非常に自然, 5=非常に不自然

表 5: 補間手法間の多重比較（Friedman + Wilcoxon, Bonferroni 補正）

評価項目	Friedman p	なし vs 従来	なし vs **提案**	従来 vs **提案**
頭部	<0.001	n.s. ($p = 0.024$)	** ($p < 0.001$)	** ($p = 0.001$)
表情	<0.001	n.s. ($p = 0.024$)	* ($p = 0.005$)	n.s. ($p = 0.024$)
全体	<0.001	* ($p = 0.007$)	*** ($p < 0.001$)	** ($p < 0.001$)

※ p は未補正値．Bonferroni 補正により有意水準を $\alpha' = 0.05/3 = 0.017$ とし，n.s.: $p \geq \alpha'$, *: $p < \alpha'$, **: $p < 0.01/3$, ***: $p < 0.001/3$

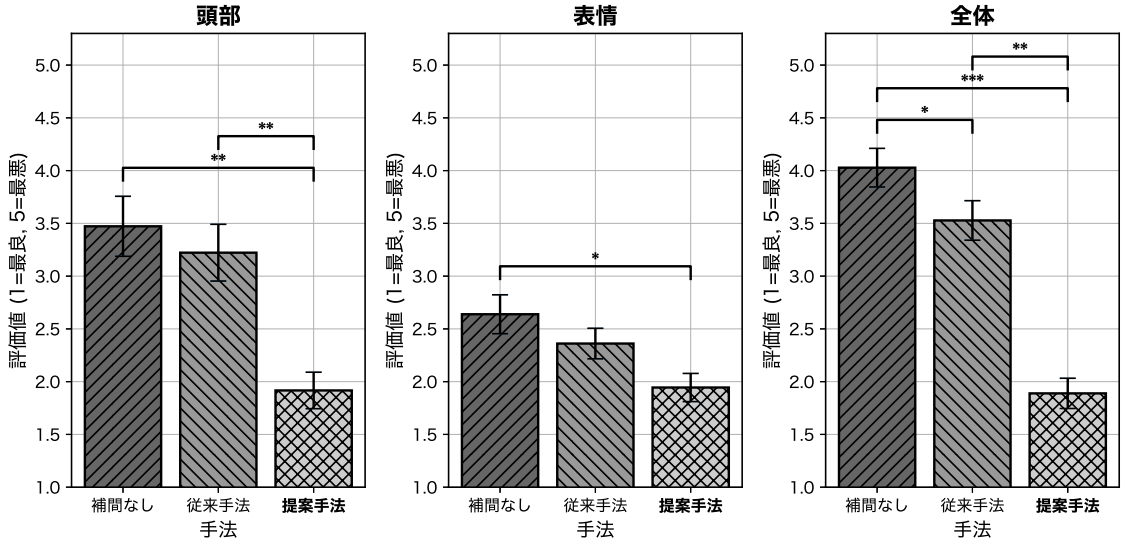

図 7: 評価項目ごとの手法比較

滑らかな補間がかえって違和感として知覚され得る．また，欠損長が長いほど欠損区間内の運動を端点情報のみで代表させることが難しくなり，T_{delay} に近い，長い欠損では軌道の妥当性が低下する可能性がある．

一方，表情では提案手法と従来手法の間に有意差は確認されなかった．表情は全体として評点が高めに推移し，手法間の差も相対的に小さかったため，提案手法と従来手法の差が統計的に検出されにくかった可能性がある．この一因として，表情変化は頭部運動に比べて画面上の変化が視覚的に小さく，欠損による表情の乱れやその補間が視聴者に知覚されにくかった可能性がある．また，表情の BlendShape 係数はスカラー値であり，頭部の位置や回転といった 3 次元の補間と比較して，従来手法の「ニュートラル状態への収束」という単純な制御でも破綻を補間しやすかった可能性がある．その結果，従来手法の段階で一定程度の破綻隠蔽が達成され，提案手法による追加的な改善量が小さくなった可能性がある．さらに，本実験においては収録の際に大きな表情変化を行わないようにしているため，表情に関しては追加条件での再検証の余地が残る．

被験者実験後の事後アンケート内の自由記述（$n = 16$）においても「表情の変化

が小さく評価が難しい」とする意見が見られた．本実験では刺激作成時に意図的にアバターの大きな表情変化を抑制しており，この実験条件が表情尺度における手法間差の検出を困難にした可能性がある．したがって，表情尺度における差が不明瞭であった点を，欠損補間手法の限界として直接的に解釈することは適切ではない．

以上の結果を元にRQ1とRQ2に回答する．

RQ1：提案手法は補間を行わない場合と比較して，視聴者が感じるアバター挙動の自然さを向上させるか？

RQ1への回答：
提案手法は補間を行わない場合と比較して，視聴者が感じるアバター挙動の自然さを向上させる．

RQ2：提案手法は従来手法と比較して，視聴者が感じるアバター挙動の自然さを向上させるか？

RQ2への回答：
提案手法は従来手法と比較して，頭部・身体運動を含む総合的なアバター挙動の自然さを向上させる一方，表情変化に対する改善効果は限定的である．

6.2 シナリオごとの主観評価

表6にシナリオ別の補間手法ごとの各動画に対する主観評価の平均値と標準偏差を，表7に手法間の差に対する検定の結果を示す．また，各シナリオ（Frame-out・Occlusion・Dynamic Occlusion・Head Rotation）について，参加者代表値の平均と標準誤差のグラフをそれぞれ図8，図9，図10，図11に示す．Friedman検定の結果，頭部および全体では，すべてのシナリオで手法間で有意差（$p < 0.001$）が認められ，事後比較では，提案手法は補間なしおよび既存手法の両方に対して全シナリオで有意に不自然さを低減した．一方，表情ではFrame-out・Occlusion・Head Rotationでは提案手法は補間なしおよび従来手法の両方に対して有意に不自然さを低減したが，Dynamic Occlusionでは，Friedman検定は有意であるものの，事後比較はいずれも補正後有意水準に至らなかった．

シナリオごとの主観評価で観測された手法間差について，各補間手法の設計上の違いに基づいて考察する．提案手法は欠損前後の短時間コンテキストから端点の運動を推定し，欠損境界での連続性を明示的に担保する設計であるため，欠損復帰時に不連続が顕在化しやすい条件ほど改善効果が大きく現れると解釈できる．ここで，不連続が顕在化しやすい条件とは，欠損長が長く，かつ欠損直前と復帰直後で姿勢・位置に差が大きい，あるいは運動方向・速度が大きく変化

表 6: シナリオ別の補間手法ごとの主観評価（平均・標準偏差）

シナリオ	手法	頭部		表情		全体	
		平均	S.D.	平均	S.D.	平均	S.D.
Frame-out	補間なし	3.44	1.25	2.47	0.65	4.06	0.92
	従来手法	3.28	1.13	2.33	0.75	3.75	1.06
	提案手法	1.61	0.90	1.69	0.71	1.69	0.93
Occlusion	補間なし	3.64	0.94	3.17	0.86	3.86	0.48
	従来手法	3.33	1.08	3.06	0.76	3.75	0.62
	提案手法	2.56	0.91	2.47	0.79	2.86	0.72
Dynamic Occlusion	補間なし	3.06	1.27	2.42	0.73	3.81	0.79
	従来手法	3.14	1.15	2.42	0.73	3.33	0.89
	提案手法	2.06	0.75	2.08	0.67	2.06	0.68
Head Rotation	補間なし	3.58	1.29	2.56	0.73	4.06	0.75
	従来手法	3.19	1.15	2.42	0.65	3.64	0.78
	提案手法	1.83	1.00	1.81	0.69	1.86	0.80

※ 1=非常に自然，5=非常に不自然

する場合である．

Frame-outにおいては，欠損直前は画面外へ向かう運動，欠損復帰直後は画面内へ戻る運動となりやすく，運動方向が逆となり，欠損境界での速度不連続が顕著になり得る．そのため，提案手法による跳躍や速度不連続の抑制がHead・Overallの改善に強く寄与した可能性がある．一方，OcclusionやDynamic Occlusionでは，遮蔽による欠損であっても，欠損前後の顔の向きや表情が比較的一定である条件も含まれ，Frame-outと比べて相対的に復帰前後の値にギャップが少なく，補間なしや従来手法でも復帰時の不連続が目立ちにくい．特にOcclusionでは，本実験刺激の設計上，演者の顔位置がほとんど変化しないように刺激動画を作成したため，欠損境界ギャップが小さく，不連続が知覚されにくかった可能性がある．その結果，提案手法による追加的な低減効果は相対的に小さく見積もられ得る．

なお，Head Rotationシナリオについては，欠損境界の不連続に加えて，刺激設計・モデル設計由来の要因が評価へ混入し得る点に留意が必要である．自由記述において，首・頭部姿勢に関する違和感への言及が多かったが，その一部は欠損補間処理そのものではなく，アバターのリギングや可動域制約といった設計要因に起因している可能性を示唆する．Head Rotationでは首を上下左右に最大90度程度傾ける動作を含んでおり，このような極端な姿勢自体が不自然に知覚された可能性を否定できない．したがって，Head Rotationの評価には，欠損補間手法の差だけでなく，姿勢表現の制約に起因する不自然さが含まれている可能性があり，欠損補間の効果が過大または過小に推定される可能性がある．したがって，シナリオ差の解釈ではそのことを前提として慎重に扱う必要がある．

表 7: シナリオ別の統補間手法間の多重比較（Friedman + Wilcoxon, Bonferroni 補正）

シナリオ	尺度	Friedman p	なし vs 従来	なし vs **提案**	従来 vs **提案**
Frame-out	頭部	<0.001	n.s. ($p = 0.221$)	** ($p < 0.001$)	** ($p < 0.001$)
	表情	<0.001	n.s. ($p = 0.272$)	** ($p = 0.002$)	* ($p = 0.011$)
	全体	<0.001	n.s. ($p = 0.061$)	*** ($p < 0.001$)	** ($p < 0.001$)
Occlusion	頭部	<0.001	* ($p = 0.013$)	*** ($p < 0.001$)	** ($p = 0.002$)
	表情	0.004	n.s. ($p = 0.396$)	* ($p = 0.005$)	* ($p = 0.009$)
	全体	<0.001	n.s. ($p = 0.271$)	** ($p < 0.001$)	** ($p < 0.001$)
Dynamic Occlusion	頭部	<0.001	n.s. ($p = 0.676$)	* ($p = 0.004$)	** ($p < 0.001$)
	表情	0.045	n.s. ($p = 0.942$)	n.s. ($p = 0.046$)	n.s. ($p = 0.047$)
	全体	<0.001	* ($p = 0.014$)	*** ($p < 0.001$)	** ($p < 0.001$)
Head Rotation	頭部	<0.001	n.s. ($p = 0.106$)	** ($p < 0.001$)	* ($p = 0.004$)
	表情	0.001	n.s. ($p = 0.222$)	** ($p = 0.002$)	** ($p = 0.003$)
	全体	<0.001	n.s. ($p = 0.017$)	*** ($p < 0.001$)	** ($p < 0.001$)

※ p は未補正値．Bonferroni 補正により有意水準を $\alpha' = 0.05/3 = 0.017$ とし，n.s.: $p \geq \alpha'$, *: $p < \alpha'$, **: $p < 0.01/3$, ***: $p < 0.001/3$

Expression で Dynamic Occlusion の差が明確にならなかった点については，移動を伴う遮蔽では注意が頭部・身体運動へ向きやすく，表情変化が主観評価に与える寄与が相対的に低下した可能性がある．また，BlendShape 係数の幾何学的に滑らかな補間は，表情の意味的変化と必ずしも一致しない可能性があるため，動きの滑らかさが人間の表情としての自然さの向上に繋がらなかった可能性がある．さらに，顔の一部のみが遮蔽される状況では，トラッキング欠損前後で BlendShape 係数に一時的なノイズが発生し，補間後の表情が不安定になった可能性がある．この不安定性により，手法による改善が一定方向に揃わず，事後比較で明確な差として確認されなかったと解釈できる．

以上の結果を元に RQ3 に回答する．

RQ3：欠損発生シナリオの違いによって，各補間手法に対する自然さの評価傾向は変化するか？

RQ3 への回答：
欠損発生シナリオの違いによって，各補間手法に対する自然さの評価傾向は変化する．特に，頭部・全体の評価ではシナリオ間で一貫した改善がみられる一方，表情の評価では欠損シナリオに依存して改善の程度が異なる．

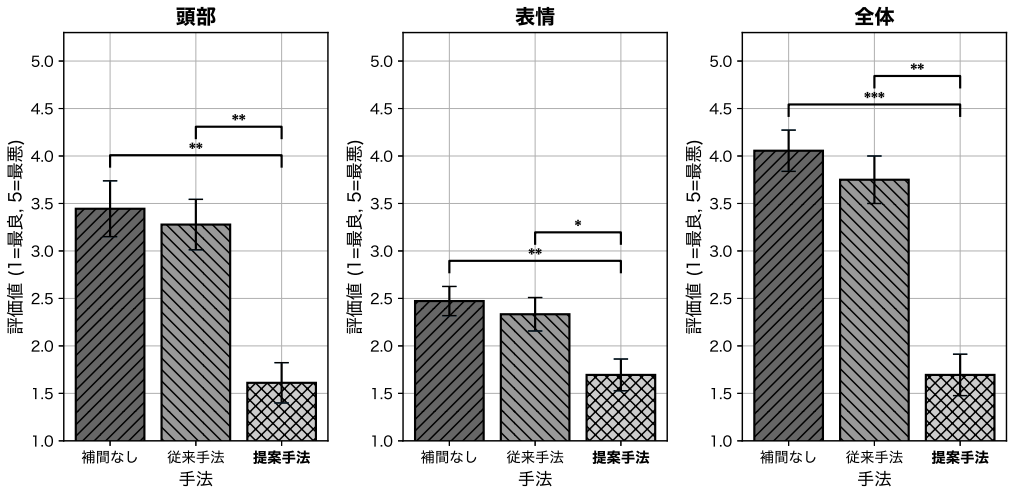

図 8: frame-out の手法比較

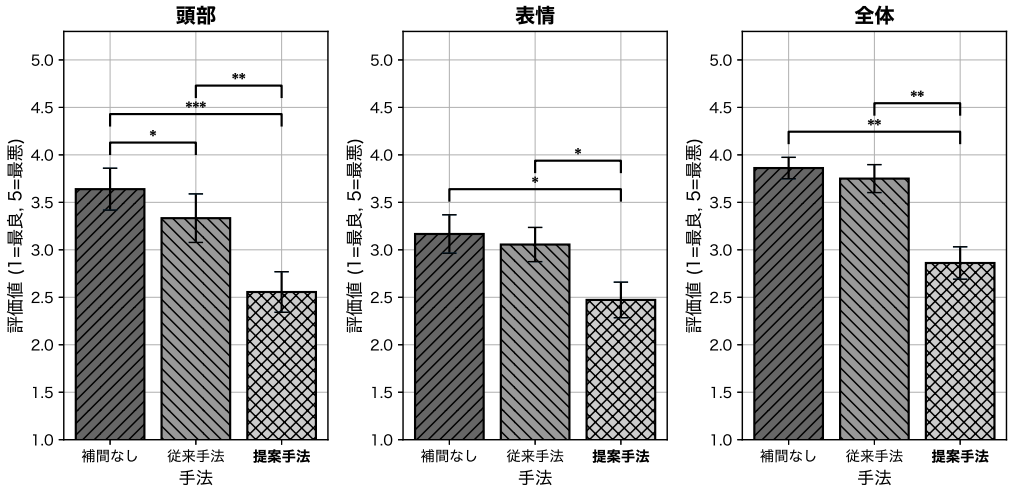

図 9: Occlusion の手法比較

6.3 事後アンケート

本節では，事後アンケート結果に基づき，参加者のVTuber視聴経験が主観評価にバイアスを与えていないかを検討する．具体的には，VTuber視聴経験が豊富な参加者はアバター挙動の微細な不連続に対して感度が高く，評価が相対的に厳しくなる可能性がある一方で，配信特有の技術的制約に慣れていることにより，不自然さを許容する傾向を示す可能性も考えられる．あわせて，自由記述から不自然さとして知覚されやすい要因を整理し，前節までの量的結果の解釈を補助する．

事後アンケートの項目「VTuberのアバターの動き・表情を普段どの程度意識して視聴していますか?」に対して「VTuberの動画を見ない」と回答した参加者をNoWatch群，それ以外「非常に意識する・少し意識する・あまり意識しない・全く意

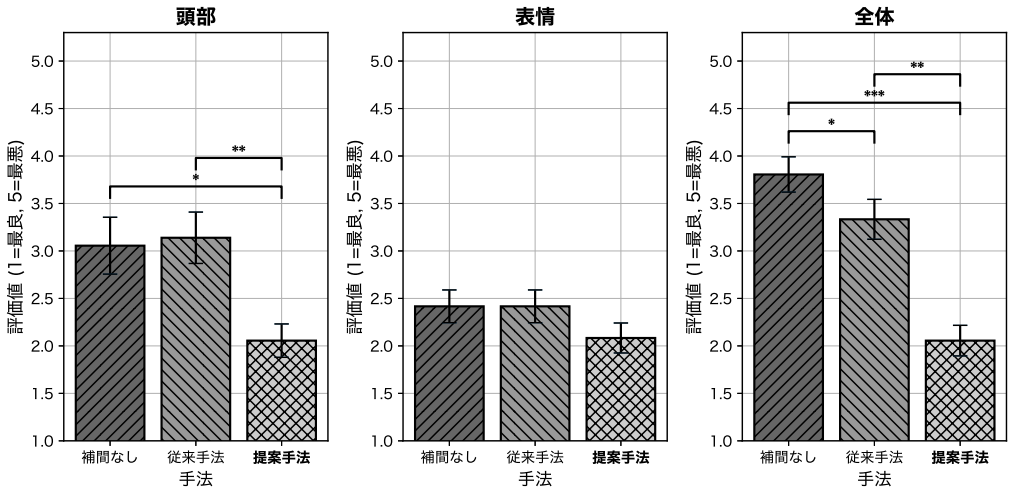

図 10: Dynamic Occlusion の手法比較

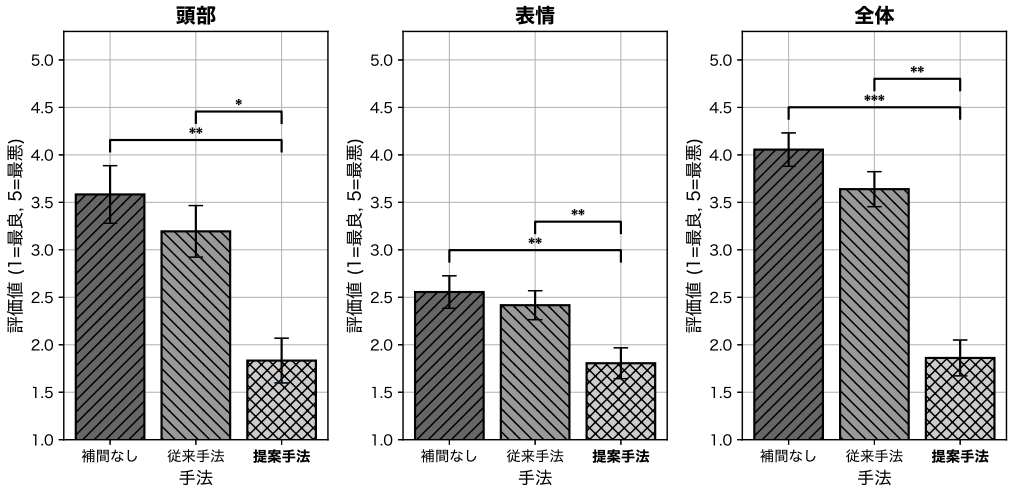

図 11: Rotation の手法比較

識しない」を回答した参加者を Watch 群として二群化した（NoWatch: $n = 7$, Watch: $n = 11$）. 視聴経験による評定傾向の違いを検討するため, 全シナリオを統合した評価結果を対象とし, 各参加者について全手法統合・補間なし・従来手法・提案手法の各条件ごとに評点を集約した代表値に基づき, NoWatch 群と Watch 群の群間差を Mann–Whitney U 検定（両側）で比較した. その結果, 群間差が比較的大きかった全手法統合の表情評価で NoWatch 群の平均値が 2.57, Watch 群が 2.09 だったが, 有意差は見られなかった（Mann–Whitney U 検定, $p = 0.091$）. 同様に, 他の評価尺度および各補間手法別の分析においても, 平均値差はいずれも 0.5 以下にとどまり, 視聴経験の有無によって評定傾向が大きく分離する状況は確認されなかった. したがって, 本実験の範囲では, VTuber 視聴経験の有無により評定が系統的に変化するという明確な証拠は得られず, 主要結果は特定の視聴経験層に依存しない可能性が示唆される. ただし, 二群化の粗さおよび群サイズの制約から, 本結果は

探索的検討として位置付ける.

自由記述 ($n = 16$) では, 不自然さとして「カクつき・カクカク」と「瞬間移動（跳躍）」を挙げる回答が多く, 特に画面端・画面外への移動や遮蔽に関連して指摘される傾向が見られた. また,「首・頭部姿勢」に関する違和感（首の可動域が大きい, 極端な上向き, 頭の傾きと首の傾きの不整合等）への言及も多かった. これらは, 前節までで議論した欠損境界での不連続や姿勢変化に関する論点と整合する傾向を示す.

以上より, VTuber視聴経験の有無による評定傾向の差は本実験では確認されず, 主要な量的結果は特定の視聴経験層に強く依存しない可能性が示唆された. また自由記述からは, カクつき・跳躍・画面端での破綻および首・頭部姿勢の不整合が不自然さの主要因として挙げられ, 表情については変化量の小ささにより評価が困難である可能性が示された.

7 おわりに

本研究はVTuber配信におけるフェイストラッキングデータの欠損がアバター動作の自然さを損なう課題に対し，リアルタイムで動作する欠損補間手法を提案した．提案手法は固定遅延バッファを用いて欠損前後の短時間コンテキストと指数重み付き移動平均によって端点の接線を推定し，各パラメータに対して3次エルミートスプラインによりC^1連続な軌道を生成する．これにより，欠損境界における位置と速度の連続性を保証し，復元誤差の最小化ではなく，視聴者が一貫した運動として受容できることを重視した補間を実現した．被験者実験では4シナリオの欠損を含む16パターンの動画を収録し，補間なし，従来手法，提案手法の3条件を適用した刺激動画を3つの尺度で評価した．実験の結果，提案手法は補間を行わない場合と比較して，頭部・表情・全体の全尺度で一貫して不自然さを低減した．また，従来手法と比較して頭部および全体において有意な改善が確認され，欠損復帰時の跳躍や速度不連続を抑制する設計の有効性が示された．一方で表情では，シナリオによっては有意な差が見られず，表情変化の評価は刺激設計や注意配分の影響を受けやすいことが示唆された．シナリオ別には，Frame-outやHead Rotationのように欠損復帰時の姿勢変化が大きく不連続が顕在化しやすい条件ほど，欠損前後の運動文脈を考慮する提案手法の効果が明確に現れた．

本稿で実施した実験では頭部および全体でアバターの動きの不自然さを減少させた一方，表情では条件により改善効果が不明瞭となった．表情評価の改善が一貫しないことは，補間手法の差に加えて，動画刺激中の表情変化量や観察者の注視（注意配分）といった要因が評定に強く影響し得ることを示す．さらに，アバターのリギングや首可動域制約など，アバターの設計要因が表情の見え方を制限し，知覚評価へ混入し得ることが示唆される．この知見を踏まえた発展的研究としては，モデル要因を統制した条件下で，表情変化を十分に含む刺激を用い，より表情に重きを置いた被験者実験を構成することが考えられる．これにより，BlendShapeの時間補間が知覚される表情の意味的変化と整合する条件としない条件とを整理できると考えられる．また，本研究の動画刺激はすべて無音であり，発話による口の動きや音声の有無が評価に与える影響については検討していない．音声の有無や声とアバターのリップシンクの整合性が自然さ評価に与える影響を検証することも，本手法の実運用上の適用範囲を整理する上で有用な研究対象となり得る．加えて，欠損長が長くなるほど欠損区間内での運動変化の未観測が増大し，端点情報だけで軌道を埋める割合が増え，短欠損（数フレーム～約1秒程度）では滑らかに接続できる一方，中～長欠損（1秒～5秒以上）では欠損中に起きた加減速や表情変化を反映できず過平滑化や復帰点との不整合が増えて補間品質が低下し得る．したがって，欠損時間の長さを独立因子として操作した分析も重要な視点となり得る．

提案手法の改善として，クォータニオン等の幾何学的に整合な表現に基づいて回転成分を補間することにより，オイラー角由来の不連続や角度折り返しに起因する不自然な回転挙動を抑制できる可能性がある．また，本研究では補間アルゴリズムとして3次エルミートスプラインを用いたが，6.2節で述べたように，BlendShape係数を幾何学的に滑らかに補間しても，表情の意味的変化としての自然さ向上には必ずしも繋がらない可能性が示唆された．例えば，表情変化における非線形性を考慮したアルゴリズムの導入や，表情パラメータ間の相関を利用した補間により，幾何学的滑らかさと知覚的自然さの乖離を縮小できる可能性がある．他にも，欠損境界でのノイズに強い欠損検出手法を導入することで，誤検出や検出遅延による補間開始，終了タイミングのずれを低減し，欠損境界での跳躍や速度不連続の発生を抑え得る．また，本稿における提案手法で用いた3秒の固定遅延は既存の配信プラットフォームが内包する遅延の範囲内では妥当である一方，視聴者との即時的なインタラクションが重視される雑談配信や，コメント反応を前提とした双方向型配信などにおいては制約となり得る．今後は，深層学習等を用いた固定遅延を前提としない軌道予測や，復帰側コンテキストを参照するハイブリッド化により，低遅延条件下でも知覚的自然さを維持可能な欠損補間へ発展させることが期待される．本研究では欠損補間を後処理として捉えたが，知覚的自然さ(Perceptual Naturalness)を高めることを目標とする品質設計問題として拡張・定式化することも発展として興味深い．文脈整合性および内的整合性（顔と身体の整合）を明示的にモデル化することにより，フェイストラッキングのみは実現できない高次の知覚的自然さを実現できる可能性がある．応用として，本手法はVTuber配信に限らず，ビデオ会議アバターやVR/AR等のリアルタイムアバター表現など，アバターを介した多様な応用領域へ展開可能である．また，本手法は特定のトラッキング対象や信号次元に依存しない時系列データの欠損補間として構成されているため，フェイストラッキングに加えて全身トラッキングの欠損補間にも適用し得る基盤技術として位置付けられる．

謝辞

被験者実験にご参加いただき，貴重なお時間を割いてご協力いただいた被験者18名の皆様に厚く御礼申し上げます. 本研究を進めるにあたり，終始ご指導・ご助言を賜りました上野秀剛准教授に深く感謝申し上げます. また，本論文の査読対応に際し，有益なご指摘とご助言を賜りました岡村真吾教授に深く感謝申し上げます.

参考文献

[1] M. Hu, M. Zhang, and Y. Wang, "Why do audiences choose to keep watching on live video streaming platforms? an explanation of dual identification framework," *Computers in Human Behavior*, Vol. 75, 2017.

[2] Z. Lu, C. Shen, J. Li, H. Shen, and D. Wigdor, "More kawaii than a real-person live streamer: Understanding how the otaku community engages with and perceives virtual youtubers," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2021.

[3] D. Kim, S. Lee, Y. Jun, Y. Shin, and J. Lee, "VTuber ' s atelier: The design space, challenges, and opportunities for VTubing," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pp. 1–23, 2025.

[4] S. Etemad, A. Arya, A. Parush, and S. Dipaola, "Perceptual validity in animation of human motion," *Computer Animation and Virtual Worlds*, Vol. 27, 2015.

[5] Y. Shuai and T. Herfet, "Towards reduced latency in adaptive live streaming," in *2018 15th IEEE Annual Consumer Communications Networking Conference (CCNC)*, pp. 1 – 4, 2018.

[6] G. Freeman, Y. Hu, R. Panchanadikar, A. L. Hall, K. Schulenberg, and L. Li, ""my audience gets to know me on a more realistic level": Exploring social vr streamers ' unique strategies to engage with their audiences," in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*," 2024.

[7] Y. J. Hwang and T. Y. Cho, "Designing for naturalness: Insights from processing fluency and visual processing," *Archives of Design Research*, Vol. 38, No. 4, pp. 129–141, 2025.

[8] X. Lei, N. Adamo-Villani, B. Benes, Z. Wang, Z. Meyer, R. E. Mayer, and A. P. Lawson, "Perceived naturalness of interpolation methods for character upper body animation." *Advances in Visual Computing. ISVC 2021. Lecture Notes in Computer Science*, Vol. 13017, 2021.

[9] J. Mezger, W. Ilg, and M. A. Giese, "Trajectory synthesis by hierarchical spatio-temporal correspondence: comparison of different methods," in *Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization*, pp. 25 – 32, 2005.

[10] P. Reitsma, J. Andrews, and N. Pollard, "Effect of character animacy and preparatory motion on perceptual magnitude of errors in ballistic motion," *Computer Graphics Forum*, Vol. 27, pp. 201 – 210, 2008.

[11] K. Mitsuo, C. Thierry, and J. K. Hodgins, "Anthropomorphism influences perception of computer-animated characters ' actions," *Social Cognitive and Affective Neuroscience*, Vol. 2, pp. 206–216, 2007.

[12] H. Welbergen, B. Van Basten, J. Egges, Z. Ruttkay, and M. Overmars, "Real time animation of virtual humans: A trade-off between naturalness and control," *Computer Graphics Forum*, Vol. 29, 2010.

[13] M. T. Tang, V. L. Zhu, and V. Popescu, "Alterecho: Loose avatar-streamer coupling for expressive vtubing," in *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 128–137, 2021.

[14] L. Chen, C. Cao, F. Torre, J. Saragih, C. Xu, and Y. Sheikh, "High-fidelity face tracking for ar/vr via deep lighting adaptation," *Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*," 2021.

[15] B. Bradwell and B. Li, "A tutorial on motion capture driven character animation," *Proceedings of the 8th IASTED International Conference on Visualization, Imaging, and Image Processing, VIIP 2008*," 2008.

[16] S. Howarth and J. Callaghan, "Quantitative assessment of the accuracy for three interpolation techniques in kinematic analysis of human movement," *Computer methods in biomechanics and biomedical engineering*, Vol. 13, pp. 847–55, 2010.

[17] R. Debski, "Real-time interpolation of streaming data," *Computer Science*, Vol. 21, 2020.