インソール型センサを用いたMLLMによる実時間スキー音声コーチング

吉原 仁 * 平野 稔祐 * Chen-Chieh Liao* Erwin Wu* 小池 英樹 *

概要. 本研究では，ウェアラブルなインソール型センサから得られる足圧と IMU データに基づき，アルペンスキーのコーチングを生成するシステムを提案する．従来のカメラ等を用いた練習支援システムは，広範囲を移動するアウトドアスポーツへの適用が困難であった．本システムは，この問題を解決しユーザが滑走中でも専門的な指導を受けられる体験を目指す．本システムは，インソール型センサの足圧と IMU のデータから英語のコーチングを生成するマルチモーダル大規模言語モデル (MLLM) と，それを自然な日本語の音声に変換するモデルの二段階で構成される．評価実験では，提案手法がテキスト類似度の評価指標で既存モデルを上回り，実際のコーチの指導に近いテキストを生成できることを確認した．

1 はじめに

スポーツの訓練において，コーチの専門知識や支援なしに，初心者が自分のスキルを向上させることは困難である．コーチングのリソースには限りがあり，練習者はしばしばプロの選手の映像を参照したり，鏡や動画を用いて自分の姿勢を比較して修正する方法に頼らざるを得ない．しかし，これらの方法は自分自身で間違いを特定し修正する必要があり，特に初心者にとっては非常に困難である．

これらの問題に対して，機械学習を活用したスポーツの練習を支援するシステムが開発されている [7, 1]．これらのシステムはユーザと熟練者の姿勢を比較して可視化したり，動画や姿勢情報からマルチモーダル大規模言語モデル (MLLM) を用いて改善方法を言語でフィードバックする．しかし，これらのシステムはカメラなどの外部装置に依存しており，広範な移動を伴うアウトドアスポーツでは適用が難しい．

そこで本論文では，ウェアラブルなインソールセンサを用いて足圧と IMU 情報からコーチングを生成するシステムを提案する（図 1）．本システムではアルペンスキーにおけるユーザの姿勢の改善に焦点を当てる．インソール型センサより得られる足圧及び IMU のデータを入力としてユーザの滑りに対するコーチングを生成する．本システムはカメラなどの外部装置に依存していない．またコーチングは実際のコーチが語りかけるような口調で行われる．これによりユーザは滑っている最中に自身の姿勢を改善して，効率よく上達することが期待される．

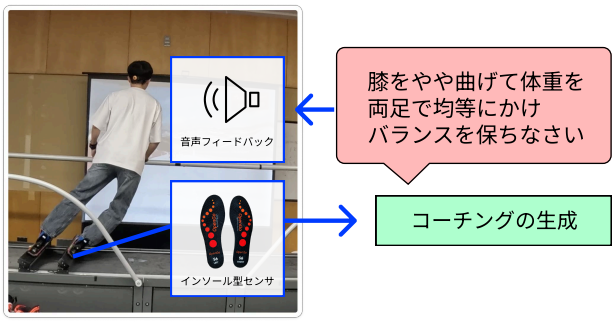

図 1. システムの概要

2 実装

本システムは足圧からコーチングを生成するコーチングモデルと，その出力をユーザに対して音声でフィードバックするモデルから構成される．

2.1 マルチモーダルスキーデータセット

システムの構築にあたって使用したデータセットについて説明する．SkiTechCoach[3] は，スキーのマルチモーダルなデータとプロによるコーチングを集めたデータセットである．データセットには 20 人の熟練者がスキーシミュレータ上でスラロームに取り組んだ際の，姿勢，足圧，コーチングのデータを含んでいる．本システムの構築にあたっては，タスクに取り組む際の被験者の姿勢，足圧のデータとそれに対するプロのコーチのコーチングコメントを用いた．

2.2 コーチング生成モデルのアーキテクチャ

コーチング生成モデルは，図 2 に示すように足圧を潜在空間に圧縮するエンコーダ層，エンコードされた足圧をテキスト空間に射影するプロジェクタ層，そしてそれらの情報からコーチングを生成する大規模言語モデルからなる．エンコーダ層には VQ-

* 東京科学大学

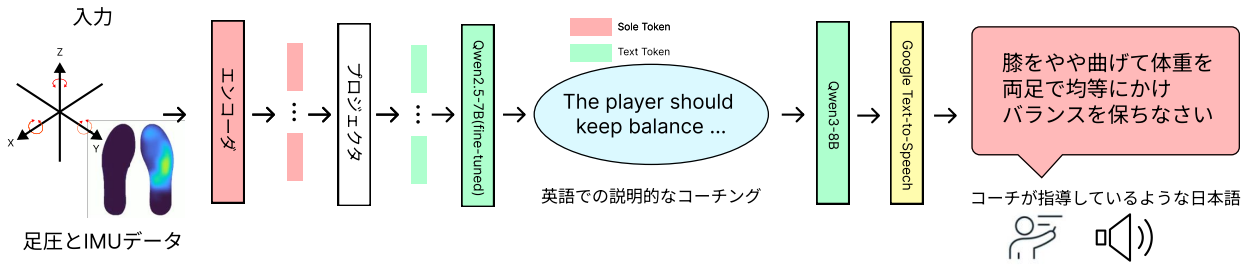

図 2. コーチング生成モデルのアーキテクチャ

VAE[9]，プロジェクタ層には多層パーセプトロン，大規模言語モデルには Qwen2.5-7B-Instruct[8] を採用した．

2.3 モデルの学習

モデルの学習は二段階で行う．第一段階として，足圧データに対してルールベースでのキャプション生成を行い，これを用いてプロジェクタ層が足圧，IMU データとテキスト空間の対応関係を捉えられるように学習をする．続く第二段階では，コーチングのデータを用いて大規模言語モデルのファインチューニングを行い，より質の高い指導を生成できるようにする．これによりモデルは，足圧と IMU データに基づくコーチング生成が可能となる．

2.4 フィードバックの手法

コーチング生成モデルにより生成されたコーチングは英語でユーザの改善点を説明的に記述したものとなっている．そのため Qwen3-8B[10] を用いて実際のコーチが指導しているかのような表現に変換する（図 2）．さらに，Google Text-to-Speech[1] を用いて音声でユーザにフィードバックする．これによってユーザは人間のコーチにより指導を受けているかのような体験をすることができる．

2.5 システムの処理時間

本システムの処理時間について，提案モデルによるコーチングの生成，そして音声合成までの一連の処理には平均で約 3 秒を要した．この処理時間は 2 枚の RTX-4090 を用いて計測された．

3 生成されたコーチングの評価

生成したコーチング文について，それに対応するコーチングの真値との類似度を BLEU-4[6]，METEOR[4]，ROUGE-L[5]，BertScore[11] を用いて評価する．表 1 中の B4 は BLEU-4，M は METEOR，R-L は ROUGE-L，BS は BertScore を意味する．Qwen2.5-VL[2] には足圧の情報をヒートマップの動画として与えた．また Qwen2.5-VL-FT は Qwen2.5-VL を提案モデルの学習に使用したデータを用いてファインチューニングしたものである．結果は表 1 に示したように全ての評価指標において提案手法が Qwen2.5-VL と Qwen2.5-VL-FT を上回った．このことから提案手法が既存のモデルよりも実際のコーチのコーチングに近いものを生成できていることが分かった．

表 1. 生成されたコーチングの評価

	B4	M	R-L	BS
Qwen2.5-VL	2.1	16.7	20.0	0.755
Qwen2.5-VL-FT	4.0	23.5	25.3	0.780
提案手法	**5.0**	**26.6**	**27.7**	**0.789**

4 議論と結論

本論文ではインソール型センサから得られた足圧と IMU のデータを用いてアルペンスキーにおけるコーチングを行うシステムを提案した．これによりカメラなどの外部装置を用いることなく，簡単に人間のコーチからレクチャーを受けているかのような体験をすることが可能となる．

今後は学習データの拡張や新たなコーチングの評価手法の提案を通して，コーチングの精度の改善に取り組む予定である．また，より使いやすいリアルタイムのコーチングシステムとするために，現在は一定の時間間隔で行われているフィードバックを，よりユーザが必要としているタイミングにすることも重要であると考えている．

謝辞

本研究は JST CRONOS JPMJCS24N8, および JST ムーンショット型研究開発事業 JPMJMS2012 の助成を受けている．

[1] https://github.com/pndurette/gTTS

参考文献

[1] K. Ashutosh, T. Nagarajan, G. Pavlakos, K. Kitani, and K. Grauman. ExpertAF: Expert Actionable Feedback from Video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13582–13594, June 2025.

[2] S. Bai, K. Chen, X. Liu, J. Wang, W. Ge, S. Song, K. Dang, P. Wang, S. Wang, J. Tang, H. Zhong, Y. Zhu, M. Yang, Z. Li, J. Wan, P. Wang, W. Ding, Z. Fu, Y. Xu, J. Ye, X. Zhang, T. Xie, Z. Cheng, H. Zhang, Z. Yang, H. Xu, and J. Lin. Qwen2.5-VL Technical Report, 2025.

[3] T. Hirano, Y. Tabei, Y. Peng, C.-C. Liao, E. Wu, and H. Koike. SkiTechCoach: A Multimodal Alpine Skiing Dataset with 3D Body Pose, Sole Pressure, and Expert Coaching. In *Proceedings of the 8th International ACM Workshop on Multimedia Content Analysis in Sports*, MMSports '25, p. 39–46, New York, NY, USA, 2025. Association for Computing Machinery.

[4] A. Lavie and A. Agarwal. Meteor: an automatic metric for MT evaluation with high levels of correlation with human judgments. In *Proceedings of the Second Workshop on Statistical Machine Translation*, StatMT '07, p. 228–231, USA, 2007. Association for Computational Linguistics.

[5] C.-Y. Lin. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out*, pp. 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics.

[6] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, p. 311–318, USA, 2002. Association for Computational Linguistics.

[7] P. Parmar, A. Gharat, and H. Rhodin. Domain Knowledge-Informed Self-supervised Representations for Workout Form Assessment. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXVIII*, p. 105–123, Berlin, Heidelberg, 2022. Springer-Verlag.

[8] Qwen, :, A. Yang, B. Yang, B. Zhang, B. Hui, B. Zheng, B. Yu, C. Li, D. Liu, F. Huang, H. Wei, H. Lin, J. Yang, J. Tu, J. Zhang, J. Yang, J. Yang, J. Zhou, J. Lin, K. Dang, K. Lu, K. Bao, K. Yang, L. Yu, M. Li, M. Xue, P. Zhang, Q. Zhu, R. Men, R. Lin, T. Li, T. Tang, T. Xia, X. Ren, X. Ren, Y. Fan, Y. Su, Y. Zhang, Y. Wan, Y. Liu, Z. Cui, Z. Zhang, and Z. Qiu. Qwen2.5 Technical Report, 2025.

[9] A. van den Oord, O. Vinyals, and K. Kavukcuoglu. Neural discrete representation learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, p. 6309–6318, Red Hook, NY, USA, 2017. Curran Associates Inc.

[10] A. Yang, A. Li, B. Yang, B. Zhang, B. Hui, B. Zheng, B. Yu, C. Gao, C. Huang, C. Lv, C. Zheng, D. Liu, F. Zhou, F. Huang, F. Hu, H. Ge, H. Wei, H. Lin, J. Tang, J. Yang, J. Tu, J. Zhang, J. Yang, J. Yang, J. Zhou, J. Zhou, J. Lin, K. Dang, K. Bao, K. Yang, L. Yu, L. Deng, M. Li, M. Xue, M. Li, P. Zhang, P. Wang, Q. Zhu, R. Men, R. Gao, S. Liu, S. Luo, T. Li, T. Tang, W. Yin, X. Ren, X. Wang, X. Zhang, X. Ren, Y. Fan, Y. Su, Y. Zhang, Y. Zhang, Y. Wan, Y. Liu, Z. Wang, Z. Cui, Z. Zhang, Z. Zhou, and Z. Qiu. Qwen3 Technical Report, 2025.

[11] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi. BERTScore: Evaluating Text Generation with BERT, 2020.