

# BikeCoach: 練習プロトコルを考慮した 自転車競技における観点別助言生成

石垣 達也<sup>1</sup> 鷲野 壮平<sup>1</sup> Matiss Rikters<sup>1</sup>  
川原田 将之<sup>1</sup> 加藤 祥太<sup>1,3</sup> 山口大貴<sup>2</sup>  
<sup>1</sup>産総研 <sup>2</sup>鹿屋体育大学 <sup>3</sup>京都大学  
ishigaki.tatsuya@aist.go.jp

## 概要

スポーツ指導の補助を指向した「助言生成」は選手の技術向上に寄与する重要な言語生成タスクである。既存の助言生成研究は練習の構成要素であるプロトコル（練習負荷設定）や助言の構造的観点を中心に考慮していない。本研究では、自転車競技を対象に、側面映像とプロトコルを入力とし、コーチング観点に基づき構造化された助言を出力するタスク「観点別助言生成 (ABAG)」を提案する。また、これを支援する大規模データセット「BikeCoach」を構築した。分析および生成実験の結果、プロトコル設定が助言内容に強く影響すること、およびLLMによる助言生成においてプロトコル情報と事例 (In-context example) の提示が品質向上に不可欠であることを明らかにした。

## 1 はじめに

大規模言語モデル (LLM) の急速な発展に伴い、スポーツ領域においても試合の要約や実況生成といったタスクでの活用が進んでいる [1, 2]。これらは主に観客の理解促進やエンゲージメント向上を目的としているが、近年では選手の技術向上を支援するための「助言生成 (Advice Generation)」が新たな研究課題として浮上している [3, 4]。

実際のトレーニング現場では、図 1 に示すように、選手は明確なプロトコル（例：「150W・90rpm で 40 秒間維持」）に従って運動を行う。コーチはその設定意図や負荷に応じて助言内容を調整する。例えば、高負荷時にはフォームの乱れを指摘し、低負荷時にはリラックスを促すといった具合である。しかし、既存の研究の多くはこうしたプロトコル変数を生成モデルの入力として考慮していない。また、生成される助言も自由記述のテキストであり、実際

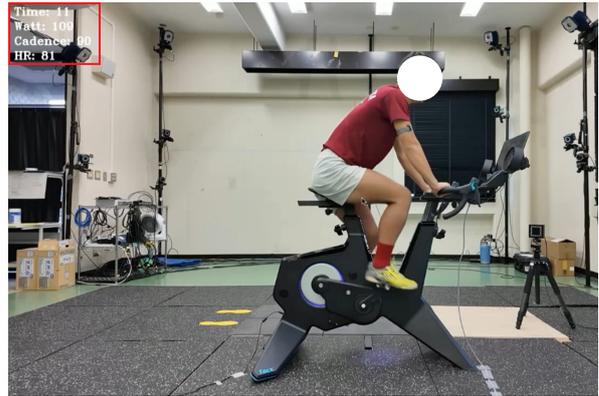


図 1: データセットの例。側面からの走行映像に加え、画面上のテレメトリ（パワー、ケイデンス等）がプロトコルとして定義される。

のコーチングで用いられる「課題の特定」「改善案の提示」といった構造化された観点 (Aspect) が欠落しているため、生成内容の分析や評価が困難であった。

自転車競技は、トレーニングの負荷がパワー (Watt) やケイデンス (rpm) といった数値で厳密に定義されるため、マルチモーダル入力（映像・数値）と言語出力（助言）の対応関係を分析する上で理想的な題材である。そこで本研究では、自転車競技を対象に以下の貢献を行う。

- 観点別助言生成 (ABAG) の提案:** 側面映像と練習プロトコルを入力とし、構造化された助言を出力するタスクを定義する。
- BikeCoach データセットの構築:** 207 名の多様な競技者による走行映像に対し、専門コーチによるフィードバックを収録・構造化したデータセットを構築した。
- プロトコルの重要性の実証:** 生成実験により、プロトコル情報と少数事例 (Few-shot) の提示が、専門家らしい助言生成に不可欠であること

を示す。

## 2 関連研究

スポーツにおける NLP: スポーツ映像のイベント検知 [5, 6] や要約生成 [7, 8] は活発に研究されている (詳細は [9, 10] を参照)。これらは「何が起きたか」を記述する記述的 (Descriptive) なタスクが主であり、近年では大規模モデルを用いたストリーミング実況生成なども提案されている [11, 12]。なお、自転車競技に特化したデータセットとしてはセンサログを中心としたもの [13] が存在するが、映像と専門家の助言を含むものはこれまでなかった。

スポーツ科学におけるコーチング: スポーツ科学の知見によれば、効果的なスキル習得には明確な課題設定 (プロトコル) と、それに対する即時のかつたを絞ったフィードバックが不可欠とされる [14, 15]。また、コーチは「技術」「機材」「戦略」といった観点ごとに助言を構造化して伝えることで、選手の理解を促している [16, 17]。運動学習におけるフィードバックの効果については Carson ら [18] も議論しており、本研究の ABAG タスクは、こうしたスポーツ科学的知見を言語生成モデルの設計に反映させる試みである。

## 3 タスク定義: 観点別助言生成

本研究では、自転車競技における観点別助言生成 (Aspect-Based Advice Generation; ABAG) を以下のように定義する。

入力: 固定カメラで撮影された側面からの走行映像  $V = \{I_t\}_{t=1}^T$  と、その区間における練習プロトコル  $P = (p, c)$  である。ここで  $p$  は目標パワー (Watt),  $c$  は目標ケイデンス (rpm) を指す。側面映像からは、選手の姿勢、上体の安定性、骨盤の動き、下肢の運動学的な特徴、ハンドル操作などが読み取れる。

出力: 表 1 に示すような、6つの観点 (Aspect) で構成される構造化された助言カードである。これらは、(1) 選手特性、(2) 選手課題、(3) 選手への推奨、(4) 機材課題、(5) 機材推奨、(6) 期待効果、から成る。これにより、単なる「頑張れ」「フォームが悪い」といった抽象的なコメントではなく、具体的かつ実行可能な (Actionable) フィードバックの生成を目指す。

表 1: 構造化された助言カードの出力例 (6つの観点)。Key-Value 形式で定義される。

観点 (Key)	値 (Value) の例
選手特性	上体は概ね安定しているが、負荷がかかると時折折みが見られる。
選手課題	設定負荷に対して肩が上がり、ハンドルへの荷重過多が見られる。
選手推奨	肩を下げ、ハンドルには手を添える程度にし、体重をペダルに乗せる。
機材課題	現在のの上体角に対してハンドル位置が少し遠い。
機材推奨	ステムを短くするかサドルを前に出し、持続可能なリーチに調整する。
期待効果	リーチを短くすることで肩がリラックスし、200W/80rpm でも骨盤が安定する。

## 4 データセット構築

ABAG タスクを支援するため、映像・プロトコル・助言を対にした大規模データセット「BikeCoach」を構築した。

### 4.1 データ収集

参加者: 体育系大学に所属する学生 207 名 (男性 171 名, 女性 36 名) が実験に参加した。彼らの専門競技は野球, サッカー, バasketボール, 陸上競技, 重量挙げ, 自転車競技など多岐にわたる。

実験環境: 計測にはスマートトレーナー (Tacx NEO Bike Plus) と Zwift を使用し、パワーとケイデンスを精密に制御・記録した。映像は DJI Osmo Pocket 3 を用い、側面から 50fps (Full HD) で記録した。実験前には、各参加者の体格に合わせてサドル高やハンドル位置の調整を行った。

プロトコル設計: 各セッションは 8 分間で構成される。ウォーミングアップとクールダウンを除くメインパート (6 分 40 秒) では、40 秒ごとに目標ケイデンス (3 段階: 低・中・高) とパワー (4 段階) が体系的に変化するよう設計された。目標パワーはトレーナーの ERG モードにより自動制御され、参加者はメトロノームに合わせて指定されたケイデンスを維持することに集中した。

### 4.2 アノテーションと構造化プロセス

収録された映像に対し、以下の手順でアノテーションを行った。

表 2: 分析に使用したコンテンツラベルの一部 (機材関連)。現象ごとの定義に基づき自動分類を行った。

IX	
ラベル ID / 名称	定義
EI_SADDLE_TOO_HIGH	膝の過伸展や骨盤の揺れにより、サドルが高すぎると判断される。
EI_SADDLE_TOO_LOW	上体が起き上がり、体幹の支持が弱くなっている。
EI_REACH_TOO_LONG	ハンドルが遠く、過度な前傾や肩の緊張が見られる。
EI_CLEAT_POSITION	クリア位置の不適合によりパワー伝達が非効率。
ER_SADDLE_HEIGHT	膝の角度や安定性を改善するために高さを調整する。
ER_HANDLE_POS	リーチやバーの高さを調整し、適切な前傾を確保する。

専門家によるフィードバック: JSPO 公認の自転車コーチ (レベル 3) が映像を確認し、40 秒ごとの区間に対してリアルタイムで口頭フィードバックを行った。コーチには、前述の 6 つの観点を意識して発話した。

文字起こし・翻訳・構造化: 収録された音声 (日本語) に対し、OpenAI の Whisper モデルを用いて文字起こしを行った。その後、GPT-4 を用いて英語への翻訳を行い、さらに GPT-4o を用いてテキストを要約し、表 1 の JSON フォーマットに構造化した。最終的に、2,058 件のデータタプルを得た。総録画時間は約 23 時間、助言テキストは約 12 万文に及ぶ。

## 5 分析: プロトコルと助言の関係

本データセットの有用性を示すため、練習プロトコル (強度設定) が助言内容にどのような影響を与えるかを定量的に分析した。

### 5.1 分析手法: コンテンツラベルの付与

分析にあたり、助言文に含まれる内容を詳細に分類するためのコンテンツラベルを設計した。データセットからランダムに抽出した 100 件の助言 (600 文) を著者が分析し、表 2 に示すようなラベルセットを定義した。その後、GPT-4o を用いてデータセット内の全助言文に対してラベル付与 (多ラベル分類) を行った。別途作成したテストセット (180 文) を用いた評価では、GPT-4o によるラベリングは適合率 94.2%、再現率 91.1% という高い精度を示し、自動ラベリングの信頼性が確認された。

### 5.2 分析結果: プロトコルの影響

ラベル分布をプロトコル条件ごとに集計したところ、明確な傾向が確認された。

ケイデンス (回転数) の影響: 機材調整に関する助言のうち、「ハンドルのリーチ調整」への言及は、高ケイデンス (110rpm) 時に 34.7% と最も高く、回転数が下がるにつれて減少した (90rpm で 31.1%、

表 3: BERTScore および ROUGE による評価 (F1)。Base: ベースライン, +Desc: プロトコル記述あり, +1-shot: さらに事例を追加。

Metric	Base	+Desc	+1-shot	Best
日本語				
BERTScore	0.756	<b>0.758</b>	0.734	+Desc
ROUGE-1	0.434	0.439	<b>0.449</b>	+1-shot
ROUGE-2	0.176	0.181	<b>0.201</b>	+1-shot
ROUGE-L	0.311	0.317	<b>0.319</b>	+1-shot
英語				
BERTScore	0.741	0.737	<b>0.746</b>	+1-shot
ROUGE-1	0.291	0.286	<b>0.303</b>	+1-shot
ROUGE-2	0.079	0.077	<b>0.086</b>	+1-shot
ROUGE-L	0.255	0.251	<b>0.264</b>	+1-shot

70rpm で 29.8%)。これは、高速回転を維持するために上体を固定しようとしてハンドルへの依存度 (荷重) が高まるためである。対照的に、「サドルの高さ調整」への言及は低ケイデンス (70rpm) 時に最も多く、高回転になるにつれて減少した (110rpm で 28.1%)。低回転ではペダルへの踏み込みトルクが大きくなるため、サドル高の不適合が膝の伸展動作に直接影響しやすいためと考えられる。

ワークロード (パワー) の影響: 機材の課題に関する助言において、「サドルが高すぎる」という指摘は、高負荷 (250W) において 35% を占め、負荷が下がるにつれて減少した (100W で 24.6%)。高負荷時にはペダリングの乱れが生じやすく、特にサドルが高すぎる場合の骨盤の揺れが顕著になるためと解釈できる。

これらの結果は、適切な助言を生成するためには、映像から見える動作だけでなく、その背景にあるプロトコル情報 (どのような負荷で走っているか) の考慮が不可欠であることを示唆している。

## 6 助言生成実験

### 6.1 実験設定

モデル: 視覚言語モデル (LVLM) として Qwen3-VL-30B-Instruct [19] を使用した。推論には A100 GPU (80GB) 上で実行した。

プロンプト設定: 以下の 3 条件を比較した。

1. **Baseline (Zero-shot):** プロトコル情報を与えず、映像のみから JSON 形式の助言を生成させる。

### Base Instruction:

Given the following data ... what advice would you give to the cyclist as a professional cycling coach ...?

Please provide more detailed explanations in the following four steps.

Step 1: Describe the characteristics of the athlete's skills.

Step 2: Point out any issues with the athlete ...

Step 3: Point out any issues with the equipment ...

Step 4: Describe the expected results ...

Output in JSON format ...

### Protocol Description (+Description):

In this training session, the athlete tries to ride a bike with the cadence of **90 rpm** with the power of **200 watts** for 40 seconds.

図 2: 実験に使用したプロンプトの構成 (抜粋). 段階的な推論 (Step-by-step) を促す指示と, プロトコル記述 (下段) から成る.

2. **+Description:** 「このセッションでは 200W/90rpm で走行している」といったプロトコル記述をプロンプトに追加する.
3. **+Description+One-shot:** プロトコル記述に加え, 別のデータから抽出した「プロトコル記述と助言のペア」を 1 例だけ文脈として提示 (In-context learning) する.

また, 生成プロンプト内には「Step 1: 選手の特徴を記述する」「Step 2: 特徴から課題を特定する」といった段階的な推論を促す指示を含め, Chain-of-Thought 的な効果も期待した構成とした.

**評価指標:** 生成された助言 (日本語および英語翻訳) に対し, 正解データ (専門家の助言) との類似度を BERTScore (F1) [20] および ROUGE (F1) [21] で評価した.

## 6.2 プロンプト設計

本実験では, モデルの推論能力を最大限に引き出すため, 図 2 に示すような段階的な指示 (Step-by-step instruction) を含むプロンプトを設計した. 具体的には, 単に助言を生成させるのではなく, (1) 動作特性の記述, (2) 選手自身の課題特定, (3) 機材の課題特定, (4) 期待される効果, という 4 つのステップを明示的に踏ませることで, 論理的整合性の取れた出力を促している. また, プロトコル記述 (Description) においては, ターゲットとなるワット数とケイデンスを具体的な数値として与えることで, モデルが「現在の負荷状況」を認識し, それに応じた適切なフィードバック (例: 高負荷ならフォーム維持を優先するなど) を選択できるよう設計した.

## 7 結果

表 3 に実験結果を示す.

**ROUGE スコアの向上:** 単にプロトコル記述を追加するだけ (+Description) では, Baseline と比較して ROUGE スコアの向上は限定的であった. しかし, 1 つの事例を提示した場合 (+Description+One-shot), 日本語・英語ともに一貫して ROUGE スコアが向上した (日本語 ROUGE-2: 0.176 → 0.201). これは, LLM が「数値 (150W)」と「助言内容」の相関をゼロショットで理解するのは困難だが, 具体的な入出力例を 1 つ提示されることで, 「高回転のときはリラックスを促す」といったプロトコルと助言の潜在的な関係性を学習できることを示唆している.

**BERTScore はスコア差を観測しない:** 一方, BERTScore の変化は微小であった. これは, 埋め込みベースの指標が意味的な類似度を評価するのに対し, ROUGE はより具体的な語彙の一致 (例: 「肩」や「骨盤」といった部位の指摘) を評価するためである. コーチングにおいては, 似たような意味でも具体的どの部位を指摘するかが重要であり, 本タスクにおいては ROUGE の向上が実質的な品質向上を反映していると考えられる.

## 8 おわりに

本研究では, 自転車競技における観点別助言生成 (ABAG) タスクを提案し, 練習プロトコルと構造化された助言を含む大規模データセット「BikeCoach」を構築した. 分析の結果, 助言内容はプロトコル (強度設定) に強く依存しており, 生成モデルにおいてもプロトコル情報と事例提示を組み合わせることが品質向上に不可欠であることを示した. 今後は, 生成された助言が実際に選手のパフォーマンス向上に寄与するかという実世界での評価や, より長期間のコンテキストを考慮した生成モデルの開発に取り組む予定である. 本研究は, 被験者を用いた実験を含んでおり, 所属機関の倫理委員会の承認を得て実施された.

## 謝辞

本研究には, 内閣府が実施する「研究開発成果の社会実装への橋渡しプログラム (BRIDGE) /AI × ロボット・サービス分野の実践的グローバル研究」により得られた成果が含まれています

## 参考文献

- [1] Yuichiro Mori, Chikara Tanaka, Aru Maekawa, Satoshi Kosugi, Tatsuya Ishigaki, Kotaro Funakoshi, Hiroya Takamura, and Manabu Okumura. Live football commentary system providing background information. In **Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (System Demonstrations)**, July 2025.
- [2] K. Tanaka, H. Nakashima, I. Noda, K. Hasida, I. Frank, and H. Matsubara. Mike: an automatic commentary system for soccer. In **Proceedings International Conference on Multi Agent Systems**, pp. 285–292, 1998.
- [3] Kumar Ashutosh, Tushar Nagarajan, Georgios Pavlakos, Kris Kitani, and Kristen Grauman. Expertaf: Expert actionable feedback from video. In **Proceedings of the Computer Vision and Pattern Recognition Conference**, pp. 13582–13594, 2025.
- [4] Tatsuki Seino, Naoki Saito, Takahiro Ogawa, Satoshi Asamizu, and Miki Haseyama. Expert comment generation considering sports skill level using a large multimodal model with video and spatial-temporal motion features. **Sensors**, Vol. 25, p. 447, 01 2025.
- [5] Silvio Giancola, Mohieddine Amine, Tarek Dghaily, and Bernard Ghanem. SoccerNet: A scalable dataset for action spotting in soccer videos. In **The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops**, June 2018.
- [6] Anthony Cioppa, Adrien Delière, Silvio Giancola, Bernard Ghanem, Marc Van Droogenbroeck, Rikke Gade, and Thomas B Moeslund. A context-aware loss function for action spotting in soccer videos. In **The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, June 2020.
- [7] Sam Wiseman, Stuart Shieber, and Alexander Rush. Challenges in data-to-document generation. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel, editors, **Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing**, pp. 2253–2263, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.
- [8] Hongmin Wang. Revisiting challenges in data-to-text generation with fact grounding. In Kees van Deemter, Chenghua Lin, and Hiroya Takamura, editors, **Proceedings of the 12th International Conference on Natural Language Generation**, pp. 311–322, Tokyo, Japan, October–November 2019. Association for Computational Linguistics.
- [9] Haotian Xia, Zhengbang Yang, Yun Zhao, Yuqing Wang, Jingxi Li, Rhys Tracy, Zhuangdi Zhu, Yuanfang Wang, Hanjie Chen, and Weining Shen. Language and multimodal models in sports: A survey of datasets and applications, 2024.
- [10] Haotian Xia, Zhengbang Yang, Junbo Zou, Rhys Tracy, Yuqing Wang, Chi Lu, Christopher Lai, Yanjun He, Xun Shao, Zhuoqing Xie, Yuanfang Wang, Weining Shen, and Hanjie Chen. SPORTU: A comprehensive sports understanding benchmark for multimodal large language models. In **The Thirteenth International Conference on Learning Representations**, 2025.
- [11] Joya Chen, Ziyun Zeng, Yiqi Lin, Wei Li, Zejun Ma, and Mike Zheng Shou. Livecc: Learning video llm with streaming speech transcription at scale. In **The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, pp. 29083–29095, 2025.
- [12] Kuangzhi Ge, Lingjun Chen, Kevin Zhang, Yulin Luo, Tianyu Shi, Liaoyuan Fan, Xiang Li, Guanqun Wang, and Shanghang Zhang. Sbench: A sports commentary benchmark for video llms, 2024.
- [13] Dorina K. Ferencsik and Erika B. Varga. Cycling activity dataset creation and application for feedback giving. **Acta Marisiensis. Seria Technologica**, Vol. 18, No. 2, pp. 29–35, 2021.
- [14] K. Anders Ericsson, Ralf T. Krampe, and Clemens Tesch-Römer. The role of deliberate practice in the acquisition of expert performance. **Psychological Review**, Vol. 100, No. 3, pp. 363–406, 1993.
- [15] Gabriele Wulf and Rebecca Lewthwaite. Optimizing performance through intrinsic motivation and attention for learning: The optimal theory of motor learning. **Psychonomic Bulletin & Review**, Vol. 23, No. 5, pp. 1382–1414, 2016.
- [16] Mark Partington and Christopher Cushion. An investigation of the practice activities and coaching behaviours of professional top-level youth soccer coaches. **Scandinavian Journal of Medicine & Science in Sports**, Vol. 23, No. 3, pp. 374–382, 2013.
- [17] Yasuhiro Yamaguchi, Kenta Sato, and Hiroshi Takahashi. A case study on skill development through structured feedback in competitive cycling. **Journal of Physical Education and Sports Pedagogy**, 2020.
- [18] Howie J. Carson and Dave Collins. The preperformance routine as a multifaceted tool: An exploratory investigation of elite golf coaches’ philosophies. **The Sport Psychologist**, Vol. 21, No. 2, pp. 170–188, 2007.
- [19] An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chuji Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiayi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report, 2025.
- [20] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. BERTScore: Evaluating text generation with bert. In **International Conference on Learning Representations**, 2020.
- [21] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In **Text Summarization Branches Out**, pp. 74–81, Barcelona, Spain, July 2004. ACL.