LEGOLAS: Learning & Enhancing Golf Skills through LLM-Augmented System

Kangbeen Ko
eyeoftyphoon@gm.gist.ac.kr
Gwangju Institute of Science and Technology
Gwangju, South Korea

Minwoo Oh
alsdn4435@gm.gist.ac.kr
Gwangju Institute of Science and Technology
Gwangju, South Korea

Minwoo Seong
seongminwoo@gm.gist.ac.kr
Gwangju Institute of Science and Technology
Gwangju, South Korea

Seungjun Kim
seungjun@gist.ac.kr
Gwangju Institute of Science and Technology
Gwangju, South Korea

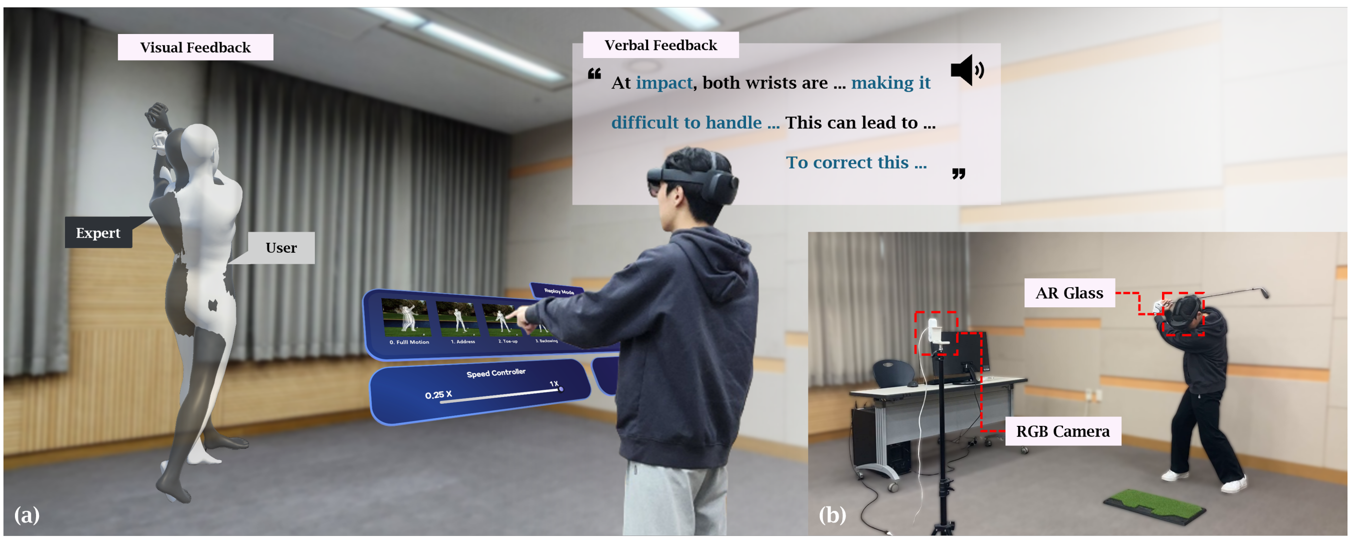

Figure 1: Overview of the system components and workflow: (a) demonstrates the combination of visual and verbal feedback, and (b) shows the experiment settings.

ABSTRACT

Effective skill acquisition in sports like golf requires both physical practice and proper feedback. Beyond error detection, valuable feedback helps learners understand underlying causes, refine mental representations, and enhance performance. While visual feedback (ViF) in self-training systems excels at identifying errors, it often lacks the capacity to address root causes or guide meaningful corrections—areas where verbal feedback (VeF) has proven highly beneficial. This study investigates the use of Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) to deliver expert-level VeF for self-training. The results show that LLM-generated VeF retains the proven advantages of traditional VeF, improving learners' mental representations and facilitating consistent progress. Additionally, integrating VeF with ViF enhances learning efficiency, self-assessment confidence, and overall performance without increasing cognitive load. This approach offers a scalable solution for effective self-training, leveraging LLMs to capture the proven advantages of VeF and bridging the gap between traditional coaching and automated systems.

CHI EA '25, April –05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/3706599.3720141

CCS CONCEPTS

• **Human-centered computing** → **Auditory feedback**; **Mixed / augmented reality**; • **Applied computing** → **Learning management systems**.

KEYWORDS

Motor skill learning, Principle-based Verbal feedback, XR training systems, Cognitive engagement

ACM Reference Format:
Kangbeen Ko, Minwoo Oh, Minwoo Seong, and Seungjun Kim. 2025. LEGOLAS: Learning & Enhancing Golf Skills through LLM-Augmented System. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (CHI EA '25).* ACM, New York, NY, USA, 10 pages. https://doi.org/3706599.3720141

1 INTRODUCTION

Sports can be interpreted as a combination of skilled movements requiring both cognitive and motor abilities. Golf, in particular, demands complex motor skills, full-body coordination, and intense mental focus [4, 43]. Acquiring effective sports techniques necessitates continuous practice and precise feedback to improve the understanding of sequential movements [2]. In the early stages of motor learning, learners concentrate on understanding movement goals and determining effective action sequences [9]. To support this process, it is critical to identify the root causes of performance issues [28]. Recent advancements in wearable sensor technology and artificial intelligence have enabled more data-driven approaches to sports training, allowing for real-time analysis and personalized feedback to enhance skill acquisition [55, 56].

Visual feedback (ViF) helps beginners, who lack a developed mental model of proper form, quickly recognize areas for correction through key visual cues. By highlighting specific points of focus, ViF enables learners to identify and address errors in their form [31]. However, despite its effectiveness in enabling easy recognition, ViF falls short in developing the core mechanics required for stable, long-term improvement in motor learning.

In contrast, principle-based verbal feedback (VeF) delivers expert-driven insights into the fundamental mechanisms behind movement discrepancies. By diagnosing fundamental problems, such as improper weight transfer, stability issues, or inadequate body coordination, learners can target core mechanisms to refine both immediate errors and grounded mechanics, ultimately developing more accurate mental models [20, 57]. Studies in sports science consistently highlight VeF's effectiveness in improving motor learning outcomes [25, 46, 58, 62]. Despite its demonstrated benefits, existing self-training systems primarily capitalize on ViF's strengths in short-term error detection, limiting their ability to provide comprehensive guidance for deeper understanding. This limitation underscores the need for feedback systems that integrate principle-based insights to support more effective learning.

To bridge this gap, we propose a self-training system for golf swings that integrates expert knowledge-based VeF with advanced ViF through XR technologies. The system leverages Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) techniques to automatically generate VeF based on expert knowledge, providing learners with personalized and actionable insights into their movements. By combining expert-informed VeF with XR-based ViF, the system offers a systematic framework for identifying and resolving discrepancies between learners' motions and target techniques, facilitating deeper understanding.

The key contributions of this study are as follows:

- Developing a system capable of delivering expert knowledge-driven VeF to enable effective self-training without the need for one-on-one coaching.
- Investigating the unique advantages of emphasized ViF and principle-based VeF in motor skill learning, emphasizing their impact on learning efficiency, confidence in self-assessment, and the enhanced outcomes achieved through their synergistic integration.

This research aims to explore possibilities of self-training systems by addressing two key research questions: First, how does LLM-generated VeF influence learning outcomes and self-confidence? Second, how do highlighted ViF and its combination with VeF impact cognitive load, performance, and user satisfaction?

2 RELATED WORKS

2.1 Mental Representation and Multi-modal Feedback in Motor-learning

Motor learning is a complex cognitive process shaped by mental representations and feedback mechanisms [2]. Mental representations serve as internal reference models, allowing learners to evaluate and refine their movements by comparing their motions against idealized versions through intrinsic feedback. These representations integrate functional, sensory, and spatiotemporal characteristics stored in long-term memory, enabling learners to execute and improve their skills [26, 41, 53]. As these mental models become more structured, learners gain the ability to enhance movement accuracy, efficiency, and proficiency independently [10, 11, 20, 27, 57].

Extrinsic feedback plays a pivotal role in shaping and refining mental models, thereby accelerating motor learning by promoting skill retention and the automation of movement patterns [34]. This feedback is especially vital for beginners, as errors made during early practice can quickly solidify into ingrained habits that are difficult to unlearn [8]. Among the various forms of extrinsic feedback, ViF has been extensively studied for its effectiveness in identifying movement discrepancies. Advanced ViF systems utilizing XR technologies offer immersive and interactive training experiences [19, 33, 47, 60]. Recent advancements in AI-driven systems have further enhanced ViF capabilities by improving the detection and emphasis of errors [24, 31]. However, despite its efficiency in recognizing errors, ViF often focuses on short-term corrections, lacking the depth required to explain the underlying principles of these corrections [51, 59].

Addressing this limitation, VeF provides a complementary approach by elucidating the core principles behind movement corrections. VeF, particularly when principle-based, delivers actionable insights into the mechanics of adjustments, emphasizing foundational elements such as stability, weight transfer, and coordination [39, 58, 62]. By focusing on the "why" and "how" of corrections, VeF fosters the development of robust mental models and accelerates the transition from conscious practice to automated movement patterns [10, 16, 25, 35, 49]. Studies have shown that VeF significantly enhances learning outcomes across a variety of sports, including sprints, tennis, and gymnastics [16, 39, 58, 62]. Furthermore, when combined with techniques like motor imagery, VeF has demonstrated an additional capacity to improve movement accuracy and consistency [51]. By addressing both the perceptual and cognitive dimensions of learning, this integrated feedback approach fosters systematic and sustained skill development.

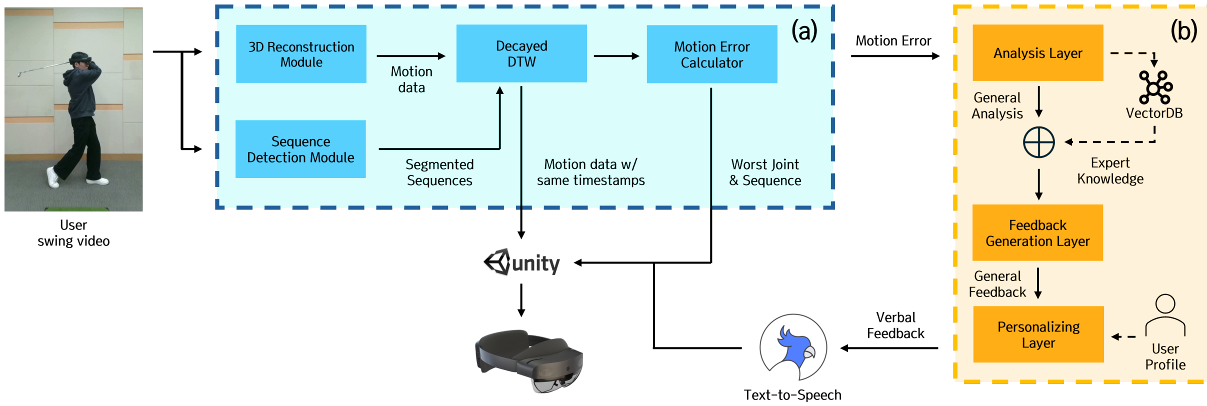

Figure 2: Representation of overall system pipeline. The system consists of two components: (a) A subsystem reconstructing the user's swing motion and calculating motion errors from RGB videos, (b) A subsystem generating VeF based on the analyzed motion using LLMs and RAG.

2.2 Automated Feedback Generation using LLMs

Recent studies on LLMs highlight their capability to provide feedback across diverse user tasks [21, 22, 50, 52]. Unlike traditional methods requiring significant time and effort to analyze individual performances, LLM-based automated feedback generation systems can deliver objective, detailed, and personalized feedback efficiently. This technology represents a transformative shift toward objective sports coaching, overcoming challenges such as human coaches' difficulty in thoroughly analyzing each learner's training data due to time constraints and limited interactions, which often results in subjective judgments, inconsistencies, and reliance on group-level feedback rather than personalized insights [36]. Furthermore, these systems enhance educational effectiveness by providing personalized coaching that reflects user characteristics and behavioral patterns [7, 15].

However, several technical limitations must be addressed for the practical application of these systems [21, 22]. A major limitation is hallucinations, in which a model generates unreliable or illogical responses [37]. In domains such as sports training where precise analysis and accurate responses are essential, such errors can compromise the reliability and usability of a system [7]. To address this, RAG technology was introduced as a method to enhance the capabilities of pre-trained language models. RAG integrates language models with an external knowledge retrieval system, enabling responses to be generated by combining the model's linguistic understanding with accurate and contextually relevant information from external sources [12, 29]. This approach significantly improves the accuracy and reliability of generated responses compared to standalone language models [12]. In the context of sports training, RAG facilitates the delivery of expert knowledge-based feedback by ensuring that coaching recommendations are grounded in dependable and well-sourced information, enhancing their overall effectiveness and trustworthiness. This approach leverages expert knowledge from authoritative sports science literature to provide precise and evidence-based feedback. By dynamically retrieving insights from specialized texts, RAG ensures accurate identification of motion errors and scientifically validated recommendations, optimizing training outcomes.

3 SYSTEM DESIGN

This study developed a system that allows users to review and replay their swing motions after execution. The user's golf swing is recorded using an ABKO APC930 QHD Webcam positioned in front of them, with the video processed by our system. Users can then review their motions overlaid on a reference model and receive feedback through augmented reality (AR). The AR scene was developed using Unity 3D and rendered to Microsoft HoloLens 2 [40] via Holographic Remoting, providing an immersive and interactive training experience.

3.1 Motion Capture & Error Calculation Module

The captured video of the user's golf swing is processed by the motion capture and analysis module (Fig. 2 (a)). This module includes two sub-components: one for detecting key swing sequences [38] and another for reconstructing 3D joint positions and quaternion data from the RGB video [30]. To align the user's motion with the target reference motion, which was selected from multiple expert demonstrations collected from the web, Decayed Dynamic Time Warping (DDTW) [19] is employed, enabling precise temporal alignment. Motion differences are analyzed by calculating vector differences in the 3D joint data for each frame, retaining directional and range information crucial for generating corrective feedback.

3.2 Verbal Feedback Generation Module

The VeF generation module (Fig. 2 (b)) processes motion errors through a structured, multi-layered architecture powered by OpenAI GPT-4 [1]. Motion error is quantified by comparing the 3D coordinates of the reference motion and the user's motion, with the hip joint as a fixed anchor point. The deviation of each joint's position is measured by computing the frame-wise 3D Euclidean distance between corresponding joints. Vector-based calculations ensure directional consistency, and an interquartile range (IQR) technique is applied to remove outliers and reduce inaccuracies

Table 1: Experimental Conditions

Condition	Visual Feedback	Verbal Feedback
BN	Basic	Non-verbal
HN	Highlighted	Non-verbal
BV	Basic	Verbal
HV	Highlighted	Verbal

Table 2: Feedback Frequency Across Blocks

Block	Feedback Frequency	Swings per Feedback	Total Feedback Instances
1	High	1	4
2	Medium	2	2
3	Low	3	1

from tracking errors. The first layer provides a general evaluation of the user's motion, using a low-temperature setting to ensure focused and deterministic results. This preliminary analysis retrieves expert knowledge from a vector database, which efficiently organizes high-dimensional data for similarity-based retrieval [13], using a RAG system [29]. It contains curated textual data sourced from reliable references [17, 23], leveraging embedding vectors to identify and extract the most relevant information. In the second layer, the retrieved knowledge is integrated with the error analysis to generate feedback tailored to each swing phase (Table 3 in Appendix), using a moderate-temperature setting to balance creativity and precision. A summary is also provided for the full motion sequence. In the final layer, the feedback is personalized to the user's profile by incorporating pre-survey data, ensuring relevance to their skill level and individual needs. A moderate-temperature setting is again applied to refine the responses, striking a balance between specificity and adaptability. The finalized feedback is then converted into audio files using a Text-to-Speech (TTS) module powered by Amazon Polly [3], ensuring the delivery of clear, reliable, and user-friendly verbal guidance.

3.3 Feedback Displayed in Augmented Reality

After the feedback generation process, users can access avatar-based ViF through AR glasses. The ViF includes two distinct 3D SMPL-based human models [32] representing the reference and the user, differentiated by color for easy identification. The movements are reconstructed using quaternion data and displayed in 3D, allowing users to view the models from any angle. This overcomes the directional limitations inherent in traditional 2D displays [44, 61]. The user interface comprises two main components: (1) a playback button for the golf swing motion and (2) a control panel (Fig. 3). The control panel features buttons labeled with images of each golf swing phase, enabling users to replay specific sequences and receive targeted feedback. Additionally, users can adjust the motion playback speed and model size using a slider bar in the control panel. In conditions that include VeF, a pause/resume button is available to manage the audio playback of VeF generated by the VeF module. The VeF is synchronized with the ViF, ensuring that verbal guidance aligns seamlessly with visual cues. To maintain clarity and focus, the VeF is designed to provide information only about the selected sequence.

4 EXPERIMENTAL DESIGN

4.1 Study Conditions

Feedback was provided through two distinct modalities: emphasized ViF, which highlights body areas with the greatest differences, and principle-based VeF, generated by LLM, using external expert

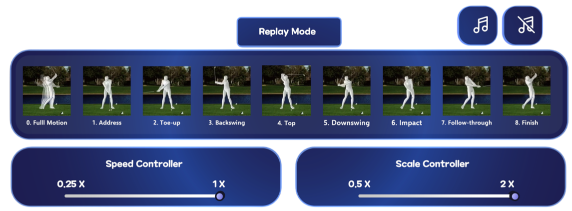

Figure 3: Interactive User Interface enabling control over motion replay phases, playback speed, model sizes, and synchronized audio feedback.

knowledge. These were organized into four experimental conditions (Table 1) to evaluate their individual and combined effects on learning efficiency and self-assessment confidence. Emphasized ViF included two variations: *B* (Basic), presenting a simple 3D model, and *H* (Highlight), emphasizing the positional differences in specific body areas. Principle-based VeF also had two categories: *N* (Non-verbal), providing only ViF, and *V* (Verbal), combining principle-based VeF with ViF.

Since the study targeted beginners to observe the formation of mental structures during the learning process. Twelve participants (age 20–27, M = 22.7, SD = 2.06; 10 males, 2 females) with less than six months of golf experience were recruited online. One participant had taken weekly golf lessons for six months. Each participant received $35 compensation. Participants completed 12 golf swing trials per condition using a 7-iron golf club, chosen for its suitability for novices. The trials were divided into three blocks of four, with feedback frequency adjusted across blocks (Table 2). The total number of feedback instances was maintained consistently across all conditions. This block design reduced variability between trials, enabled the analysis of learning patterns over time, and enhanced the reliability of statistical results through data aggregation [54].

The experiment was counterbalanced using a Latin square design to minimize learning effects. After each trial, participants provided a subjective self-assessment of their performance and rated their confidence in this evaluation, based on prior research on feedback withdrawal and its impact on error estimation confidence [54]. By engaging in self-evaluation prior to receiving external feedback, participants were able to enhance both their technical skills and their ability to assess their own performance more effectively [48].

4.2 Collected Data

This study investigated whether VeF generated by a LLM could facilitate effective training. The impact of VeF on learning effectiveness, cognitive load, and self-evaluation confidence was assessed using the following metrics:

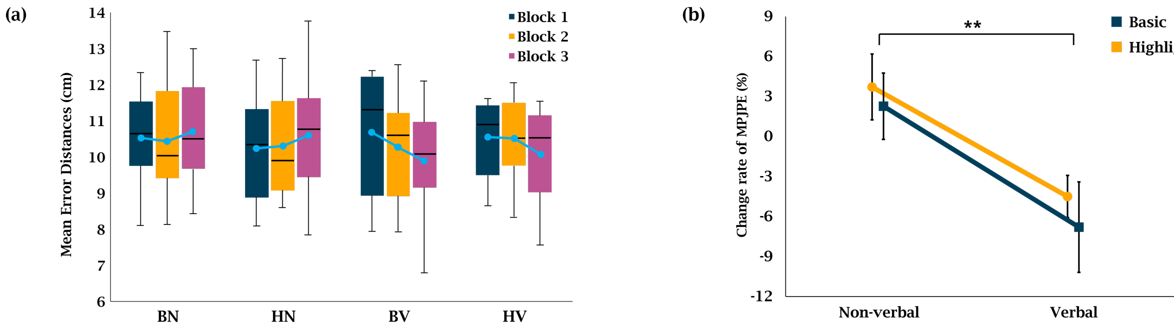

Figure 4: (a) Mean error distances (cm) across conditions and blocks, with the averages for each block connected by a blue line to visualize the trend. (b) The change rate of MPJPE (%) from blocks 1 to 3. Significance levels are indicated: $^{**}p < 0.01$

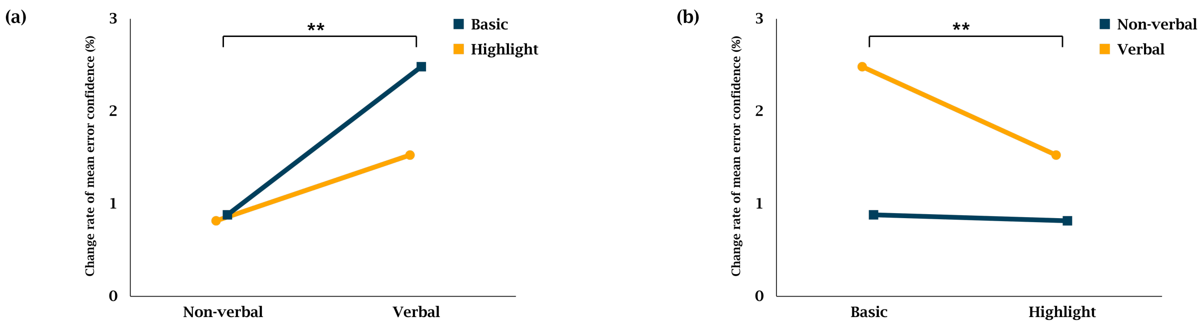

Figure 5: (a) The change rate of mean error confidence (%) in each trial across Non-verbal and Verbal conditions, (b) The change rate of mean error confidence (%) in each trial across Basic and Highlight conditions. Significance levels are indicated: $^{**}p < 0.01$

(1) **Mean Per Joint Position Error (MPJPE):** Movement accuracy was measured using MPJPE, which calculates the average positional error between target and participant joint positions. MPJPE is a reliable metric widely used in 3D pose estimation [6, 31, 64]. It was calculated with reconstructed 3D models afterward.
(2) **NASA-TLX:** Cognitive load was assessed using the NASA-TLX survey, which evaluates mental, physical, and temporal demands, as well as performance, effort, and frustration. This tool is extensively applied in motor learning and sports research [14, 42]. Surveys were administered immediately after each condition.
(3) **Self-evaluation Confidence:** Participants rated their confidence in evaluating their own performance using a 1- to 5-point Likert scale, after each trial based on previous research that suggests cognitive structure development enhances self-evaluation confidence [54].

5 RESULTS & DISCUSSIONS

This study employed a Two-way Repeated Measures ANOVA (RM ANOVA) to analyze the experimental results and examine the effects of two independent variables, ViF and VeF, on the dependent variables; MPJPE, NASA-TLX, and Error Confidence. As no significant interaction effects were observed between the ViF and VeF, the analysis focused on the main effects. Data normality was assessed using the Shapiro-Wilk test, and for variables violating normality—namely the frustration subscale of NASA-TLX and the confidence slope—Conover's test was applied as a non-parametric alternative. Qualitative insights were obtained from user interview data.

5.1 Quantitative Data

Learning performance was evaluated using MPJPE values, with average values calculated for each block to analyze error trends. No significant differences were observed between conditions in blocks 1 and 2, where feedback frequency was relatively high. However, in block 3, where feedback frequency was relatively low, errors increased under non-VeF conditions (*BN*, *HN*) but continued to decrease under VeF conditions (*BV*, *HV*) (Fig. 4(a)). These findings suggest that LLM-generated VeF can reduce reliance on external feedback, mirroring the effects of human-delivered VeF demonstrated in prior research [16, 25, 49, 58, 62]. Additionally, RM ANOVA showed that VeF significantly improved performance ($^{**}p = 0.006$) (Fig. 4(b)), in contrast to ViF, which did not demonstrate a significant effect. This highlights the distinct contribution of VeF to performance enhancement, particularly in reducing errors under low-feedback conditions.

Confidence slope analysis further supported this finding, showing that participants under the VeF condition exhibited a significantly higher increase in confidence levels over the 12 trials than those

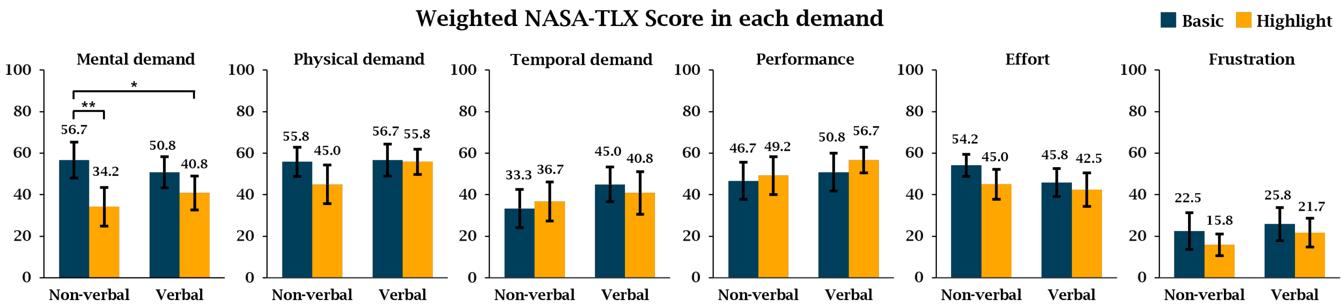

Figure 6: Cognitive workload (NASA-TLX) for each dimension across Non-verbal and Verbal condition. Significance levels are indicated: * $p < 0.05$, ** $p < 0.01$

under the non-VeF condition (p = 0.001) (Fig. 5(a)). The change rate of error confidence was calculated by measuring the rate of change in the normalized confidence score from the start to the end point for each user and condition. These results reinforce the MPJPE findings and emphasize the critical role of VeF in enhancing independent learning. However, confidence increased significantly less in the highlighted ViF condition compared to the basic ViF condition(p = 0.008) (Fig. 5(b)). This may be due to immediate ViF in the highlighted ViF condition hindering confidence formation by presenting discrepancies with self-assessment [45]. In contrast, the basic ViF condition allowed participants more time for reflection, enabling motor memory stabilization and faster confidence development [63].

The NASA-TLX analysis showed significant differences in cognitive workload across conditions, particularly in the Mental Demand category (Fig. 6). RM ANOVA indicated that highlighted ViF significantly reduced mental demand (p < 0.001). Post-hoc tests revealed that the highlighted Vif without VeF (*HN*) condition imposed significantly lower mental demand than basic ViF without VeF (*BN*) (p = 0.006), and highlighted ViF with VeF (*HV*) was also lower than basic ViF without VeF (*BN*) (p = 0.029). While highlighted ViF without VeF (*HN*) showed lower cognitive workload than highlighted ViF with VeF (*HV*), the difference was not statistically significant. Although the combined condition (*HV*) did not significantly reduce workload compared to the highlighted ViF-only condition (*HN*), the absence of a significant increase is noteworthy. This suggests that the combination of VeF and ViF enhances performance without adding cognitive strain, demonstrating its effectiveness in supporting skill acquisition and self-directed learning.

5.2 Qualitative Data

The user interviews explored the effects of highlighted ViF and LLM-generated VeF on golf learning. Participants noted that highlighted ViF (*H*) helped them identify differences between their movements and the target movements *(P3, P8, P9)*. However, some felt that focusing on specific body parts alone was insufficient for fully correcting movements *(P4, P6, P11)*. While a few participants mentioned that processing both types of feedback simultaneously was somewhat overwhelming *(P9, P10)*, most participants, representing over half, found the combined condition (*HV*) to be the most effective. This supports quantitative findings that combining VeF and ViF does not significantly increase cognitive workload while still enhancing overall effectiveness. Participants suggested adding numerical feedback for movement errors *(P1, P11)* and enabling real-time comparisons with target movements *(P4, P12)*. Overall, the interviews highlight the value of VeF in fostering deeper understanding, reducing perceived difficulty, and enhancing the learning experience.

6 CONCLUSIONS

This study demonstrated that the benefits of VeF, previously established in sports training research, can be effectively replicated using LLM in motor-skill learning. The results showed that LLM-generated VeF significantly improved movement accuracy while reducing reliance on external feedback. Additionally, the combined condition (*HV*) enhanced both performance and user satisfaction without increasing cognitive load, underscoring the complementary strengths of VeF and ViF. VeF provided precise, principle-based movement corrections, while ViF offered intuitive and immediate error detection, together achieving results comparable to traditional coaching.

These findings suggest that LLM-generated VeF can be integrated into sports self-training systems. When paired with ViF, these systems can deliver autonomous, user-friendly training solutions tailored to individual needs. Advances in LLM technology, especially with multimodal-LLMs, promise further improvements in personalization, accuracy, and accessibility, thereby democratizing high-quality feedback while reducing dependence on costly traditional methods. Real-time tactile cues, enabled by advancements in haptic feedback technology, can further enhance motor learning and movement execution in sports training system [18].

Despite the promising results, some limitations were identified, such as motion tracking challenges caused by occlusion during the Finish stage and the lack of detailed biomechanical feedback. Moreover, although this study showed short-term improvements in motor skill acquisition, its long-term effectiveness without continuous feedback remains uncertain. Future research should aim to improve motion capture accuracy and incorporate real-time feedback to correct errors instantly, ensuring more effective and responsive learner support. Longitudinal studies are also needed to assess whether the observed performance gains persist over time and to explore strategies for sustaining skill acquisition.

This study highlights the potential of combining LLM-generated VeF with ViF in self-training systems for sports, especially for golf. By demonstrating the scalability and effectiveness of AI-driven feedback, it provides a framework for developing personalized, adaptive training solutions that enable individuals to achieve professional-level skill refinement independently.

ACKNOWLEDGMENTS

This work was supported by the GIST-MIT Research Collaboration grant funded by the GIST in 2025, the 'Project for science and technology opens the future of the region' program through the Innopolis Foundation funded by Ministry of Science and ICT (Project Number: 2025-DD-UP-0312), and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00343397).

REFERENCES

[1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

[2] Fran Allard and Janet L Starkes. 1991. Motor-skill experts in sports, dance, and other domains. (1991).

[3] Amazon Web Services, Inc. 2025. *Amazon Polly Documentation.* Amazon Web Services, Inc. https://docs.aws.amazon.com/polly/ Accessed: 2025-01-21.

[4] Philipp Barzyk and Markus Gruber. 2024. Motor learning in golf—a systematic review. *Frontiers in Sports and Active Living* 6 (2024), 1324615.

[5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[6] Yun-Wei Chiu, Kuei-Ting Huang, Yuh-Renn Wu, Jyh-How Huang, Wei-Li Hsu, and Pei-Yuan Wu. 2024. 3D Baseball Pitcher Pose Reconstruction Using Joint-wise Volumetric Triangulation and Baseball Customized Filter System. *IEEE Access* (2024).

[7] Mark Connor and Michael O'Neill. 2023. Large language models in sport science & medicine: Opportunities, risks and considerations. *arXiv preprint arXiv:2305.03851* (2023).

[8] Kevin W Eva and Glenn Regehr. 2008. "I'll never play professional football" and other fallacies of self-assessment. *Journal of Continuing Education in the Health Professions* 28, 1 (2008), 14–19.

[9] PM Fitts. 1967. Human performance. *Brooks/Cole* (1967).

[10] Cornelia Frank, William M Land, Carmen Popp, and Thomas Schack. 2014. Mental representation and mental practice: experimental investigation on the functional links between motor memory and motor imagery. *PloS one* 9, 4 (2014), e95175.

[11] Cornelia Frank, William M Land, and Thomas Schack. 2016. Perceptual-cognitive changes during motor learning: the influence of mental and physical practice on mental representation, gaze behavior, and performance of a complex action. *Frontiers in psychology* 6 (2016), 1981.

[12] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, and Haofen Wang. 2023. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997* (2023).

[13] Yikun Han, Chunjiang Liu, and Pengfei Wang. 2023. A comprehensive survey on vector database: Storage and retrieval technique, challenge. *arXiv preprint arXiv:2310.11703* (2023).

[14] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[15] Narayan Hegde, Madhurima Vardhan, Deepak Nathani, Emily Rosenzweig, Cathy Speed, Alan Karthikesalingam, and Martin Seneviratne. 2024. Infusing behavior science into large language models for activity coaching. *PLOS Digital Health* 3, 4 (2024), e0000431.

[16] Souhail Hermassi, Maha Sellami, El Ghali Bouhafs, René Schwesig, and Andrea De Giorgio. 2019. Effect of verbal instruction on motor learning ability of anaerobic and explosive exercises in physical education university students. *Frontiers in psychology* 10 (2019), 2097.

[17] Ben Hogan. 2024. *Ben Hogan's five lessons: The modern fundamentals of golf.* Simon and Schuster.

[18] Seokhyun Hwang, Jeongseok Oh, Seongjun Kang, Minwoo Seong, Ahmed Ibrahim Ahmed Mohamed Elsharkawy, and SeungJun Kim. 2024. Ergopulse: Electrifying your lower body with biomechanical simulation-based electrical muscle stimulation haptic system in virtual reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–21.

[19] Atsuki Ikeda, Dong-Hyun Hwang, Hideki Koike, Gerd Bruder, Shunsuke Yoshimoto, and Sue Cobb. 2018. AR based Self-sports Learning System using Decayed Dynamic TimeWarping Algorithm.. In *ICAT-EGVE*. 171–174.

[20] Edmund Jacobson. 1932. Electrophysiology of mental activities. *The American Journal of Psychology* 44, 4 (1932), 677–694.

[21] Qinjin Jia, Jialin Cui, Haoze Du, Parvez Rashid, Ruijie Xi, Ruochi Li, and Edward Gehringer. 2024. LLM-generated Feedback in Real Classes and Beyond: Perspectives from Students and Instructors. In *Proceedings of the 17th International Conference on Educational Data Mining*. 862–867.

[22] Qinjin Jia, Jialin Cui, Ruijie Xi, Chengyuan Liu, Parvez Rashid, Ruochi Li, and Edward Gehringer. 2024. On Assessing the Faithfulness of LLM-generated Feedback on Student Assignments. In *Proceedings of the 17th International Conference on Educational Data Mining*. 491–499.

[23] Theodore P Jorgensen. 2013. *The physics of golf.* Springer Science & Business Media.

[24] Chan-Yang Ju, Jong-Hyeon Kim, and Dong-Ho Lee. 2023. GolfMate: Enhanced Golf Swing Analysis Tool through Pose Refinement Network and Explainable Golf Swing Embedding for Self-Training. *Applied Sciences* 13, 20 (2023), 11227.

[25] Ryuta Kawashima, Nobumoto Tajima, Hajime Yoshida, Katsuo Okita, Takeo Sasaki, Torsten Schormann, Akira Ogawa, Hiroshi Fukuda, and Karl Zilles. 2000. The effect of verbal feedback on motor learning—a PET study. *Neuroimage* 12, 6 (2000), 698–706.

[26] Taeho Kim, Cornelia Frank, and Thomas Schack. 2017. A systematic investigation of the effect of action observation training and motor imagery training on the development of mental representation structure and skill performance. *Frontiers in human neuroscience* 11 (2017), 499.

[27] Taeho Kim, Hyunsik Park, and Thomas Schack. 2019. A functional link between mental representation in long-term memory and cognitive performance in working memory. *Advances in Cognitive Psychology* 15, 1 (2019), 11.

[28] Kristan A Leech, Ryan T Roemmich, James Gordon, Darcy S Reisman, and Kendra M Cherry-Allen. 2022. Updates in motor learning: implications for physical therapist practice and education. *Physical therapy* 102, 1 (2022), pzab250.

[29] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems* 33 (2020), 9459–9474.

[30] Jiefeng Li, Chao Xu, Zhicun Chen, Siyuan Bian, Lixin Yang, and Cewu Lu. 2021. Hybrik: A hybrid analytical-neural inverse kinematics solution for 3d human pose and shape estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 3383–3393.

[31] Chen-Chieh Liao, Dong-Hyun Hwang, Erwin Wu, and Hideki Koike. 2023. Ai coach: A motor skill training system using motion discrepancy detection. In *Proceedings of the Augmented Humans International Conference 2023*. 179–189.

[32] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. 2023. SMPL: A skinned multi-person linear model. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*. 851–866.

[33] Dizhi Ma, Xiyun Hu, Jingyu Shi, Mayank Patel, Rahul Jain, Ziyi Liu, Zhengzhe Zhu, and Karthik Ramani. 2024. avaTTAR: Table Tennis Stroke Training with Embodied and Detached Visualization in Augmented Reality. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–16.

[34] T Mahmood and A Darzi. 2004. The learning curve for a colonoscopy simulator in the absence of any feedback: no feedback, no learning. *Surgical Endoscopy And Other Interventional Techniques* 18 (2004), 1224–1230.

[35] Nogal Marta. 2020. Effectiveness of verbal feedback in complex motor skill learning. *Theory and Practice of Physical Culture* 8 (2020), 63–66.

[36] Robert J Mason, Damian Farrow, and John AC Hattie. 2020. Sports coaches' knowledge and beliefs about the provision, reception, and evaluation of verbal feedback. *Frontiers in Psychology* 11 (2020), 571552.

[37] Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. On faithfulness and factuality in abstractive summarization. *arXiv preprint arXiv:2005.00661* (2020).

[38] William McNally, Kanav Vats, Tyler Pinto, Chris Dulhanty, John McPhee, and Alexander Wong. 2019. Golfdb: A video database for golf swing sequencing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 0–0.

[39] Christopher Meier, Cornelia Frank, Bernd Gröben, and Thomas Schack. 2020. Verbal instructions and motor learning: How analogy and explicit instructions influence the development of mental representations and tennis serve performance. *Frontiers in Psychology* 11 (2020), 2.

[40] Microsoft. 2025. *HoloLens 2*. Retrieved Accessed: Jan 16, 2025 from https://learn.microsoft.com/en-us/hololens/

[41] Cristina Elena Moraru, Emilia Florina Grosu, Liliana Elisabeta Radu, and Rares Puni. 2014. Influence of mental training on the execution technique in rhythmic gymnastics. In *3rd WORLD CONFERENCE on DESIGN, ARTS AND EDUCATION*.

[42] T Mullen et al. 2021. Influence of Contextual Factors, Technical Perform... *ijspp* 2019 (2021), 0998.

[43] Steven M Nesbit and Monika Serrano. 2005. Work and power analysis of the golf swing. *Journal of sports science & medicine* 4, 4 (2005), 520.

[44] Masaki Oshita, Takumi Inao, Tomohiko Mukai, and Shigeru Kuriyama. 2018. Self-training system for tennis shots with motion feature assessment and visualization. In *2018 International Conference on Cyberworlds (CW)*. IEEE, 82–89.

[45] Ori Ossmy and Roy Mukamel. 2018. Behavioral and neural effects of congruency of visual feedback during short-term motor learning. *NeuroImage* 172 (2018), 864–873.

[46] Carmen Pârvu, George Dănuț Mocanu, Petronel Cristian Moisescu, and Zukhro Bahadirovna Khamraeva. 2024. The Optimisation of Education through Feedback and Advanced Technology in Motor Learning and the Correction of Technical Sports Errors. *Research and education* 10 (2024), 2559–2033.

[47] Stefan Pastel, Katharina Petri, Chien-Hsi Chen, Ana Milena Wiegand Cáceres, Meike Stirnatis, Carlo Nübel, Lasse Schlotter, and Kerstin Witte. 2023. Training in virtual reality enables learning of a complex sports movement. *Virtual Reality* 27, 2 (2023), 523–540.

[48] Jae T Patterson, Matthew McRae, and Sharon Lai. 2016. Accuracy of subjective performance appraisal is not modulated by the method used by the learner during motor skill acquisition. *Perceptual and Motor Skills* 122, 2 (2016), 650–665.

[49] Nicole K Rendos, Laura Zajac-Cox, Rahul Thomas, Sumire Sato, Steven Eicholtz, and Trisha M Kesar. 2021. Verbal feedback enhances motor learning during post-stroke gait retraining. *Topics in stroke rehabilitation* 28, 5 (2021), 362–377.

[50] Sara Riazi and Pedram Rooshenas. 2024. LLM-Driven Feedback for Enhancing Conceptual Design Learning in Database Systems Courses. *arXiv preprint arXiv:2412.17892* (2024).

[51] Nicolas Robin, Eric Joblet, Emmanuel Roublot, and Guillaume R Coudevylle. 2020. The beneficial effect of combining feedback, observational learning and motor imagery on football pass performance. *Motricidade* 16, 1 (2020), 55–65.

[52] Theresa Ruwe and Elisabeth Mayweg-Paus. 2024. Embracing LLM Feedback: the role of feedback providers and provider information for feedback effectiveness. In *Frontiers in Education*, Vol. 9. Frontiers Media SA, 1461362.

[53] Thomas Schack, Kai Essig, Cornelia Frank, and Dirk Koester. 2014. Mental representation and motor imagery training. *Frontiers in human neuroscience* 8 (2014), 328.

[54] Richard A Schmidt and Janice L White. 1972. Evidence for an error detection mechanism in motor skills: A test of Adams' closed-loop theory. *Journal of Motor Behavior* 4, 3 (1972), 143–153.

[55] Minwoo Seong, Gwangbin Kim, Yumin Kang, Junhyuk Jang, Joseph DelPreto, and SeungJun Kim. 2024. Counterfactual Explanation-Based Badminton Motion Guidance Generation Using Wearable Sensors. *arXiv preprint arXiv:2405.11802* (2024).

[56] Minwoo Seong, Gwangbin Kim, Dohyeon Yeo, Yumin Kang, Heesan Yang, Joseph DelPreto, Wojciech Matusik, Daniela Rus, and SeungJun Kim. 2024. Multisense-badminton: Wearable sensor–based biomechanical dataset for evaluation of badminton performance. *Scientific Data* 11, 1 (2024), 343.

[57] Bianca A Simonsmeier and Susanne Buecker. 2017. Interrelations of imagery use, imagery ability, and performance in young athletes. *Journal of Applied Sport Psychology* 29, 1 (2017), 32–43.

[58] Marcin Starzak, Michał Biegajło, Marta Nogal, Tomasz Niźnikowski, Tadeusz Ambroży, Łukasz Rydzik, and Jarosław Jaszczur-Nowicki. 2022. The role of verbal feedback in the motor learning of gymnastic skills: A systematic review. *Applied Sciences* 12, 12 (2022), 5940.

[59] Diane M Ste-Marie, Barbi Law, Amanda M Rymal, O Jenny, Craig Hall, and Penny McCullagh. 2012. Observation interventions for motor skill learning and performance: an applied model for the use of observation. *International Review of Sport and Exercise Psychology* 5, 2 (2012), 145–176.

[60] Kosuke Takahashi, Dan Mikami, Mariko Isogawa, Yoshinori Kusachi, and Naoki Saijo. 2019. Vr-based batter training system with motion sensing and performance visualization. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 1353–1354.

[61] Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2013. MotionMA: Motion modelling and analysis by demonstration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1309–1318.

[62] Sameh Wali-Menzli, Sarra Hammoudi-Nassib, Souhaila Ismail, Sabra Riahi Hammoudi, Ines Knani Hamrouni, and Mohamed Jarraya. 2019. Role of the mental representation in enhancing motor learning and performing gymnastic element. *Science of Gymnastics Journal* 11, 1 (2019), 91–102.

[63] Dangxiao Wang, Teng Li, Gaofeng Yang, and Yuru Zhang. 2017. Effects of concurrent and delayed visual feedback on motor memory consolidation. *IEEE transactions on haptics* 10, 3 (2017), 350–357.

[64] Jinbao Wang, Shujie Tan, Xiantong Zhen, Shuo Xu, Feng Zheng, Zhenyu He, and Ling Shao. 2021. Deep 3D human pose estimation: A review. *Computer Vision and Image Understanding* 210 (2021), 103225.

A VERBAL FEEDBACK EXAMPLES

This table provides examples of VeF (Verbal Feedback) given at each stage. VeF focuses on three key elements: (1) identifying the error's position, (2) explaining the problem caused by the error, and (3) offering guidance on how to resolve it. Since the VeFs were delivered as audio, the prompts were crafted for brevity and clarity, using concise sentences to promote effective understanding and implementation.

B THEMATIC ANALYSIS OF USER INTERVIEWS

This appendix presents a thematic analysis of user interviews on the effectiveness of feedback modalities in a golf swing learning system. The analysis identifies key themes across participants (P1–P12) and summarizes consistent insights.

Following *Braun and Clarke's* six-phase framework[5], a widely recognized qualitative method, the analysis systematically proceeded through: **(1) Data familiarization**, **(2) Initial coding**, **(3) Theme identification**, **(4) Theme refinement**, **(5) Theme definition**, and **(6) Report synthesis**. Themes were iteratively refined through cross-referencing participant responses to ensure clarity and coherence.

B.1 Most Challenging Feedback Modality

- BN (Basic ViF Only) was the most difficult: Users found it hard to recognize mistakes and determine how to correct them. (P1, P2, P3, P4, P6, P8, P10, P12)
- HN (Highlighted ViF Only) lacked guidance: While highlights indicated problem areas, they did not explain how to fix them. (P2, P3, P4, P6, P10, P12)
- BV (Basic ViF + VeF) was more helpful: VeF provided clear explanations, making corrections easier. (P1, P2, P4, P5, P6, P8, P9, P10, P12)
- HV (Highlighted ViF + VeF) had mixed reviews: Some users found it beneficial, while others experienced information overload. (P3, P4, P6, P8, P10, P11)

B.2 Effectiveness of Highlighted ViF

- Helped users identify motion differences between themselves and the expert model. (P1, P2, P4, P5, P6, P9, P10)
- Effective for focusing on specific areas, but less useful for full-body corrections. (P3, P6, P8, P9, P12)

B.3 Effectiveness of VeF

- Crucial for understanding corrections: Provided detailed explanations of mistakes and solutions. (P1, P3, P5, P7, P8, P10, P11)
- More useful than highlights alone: Helped users plan adjustments rather than just recognize mistakes. (P1, P4, P6, P8, P10, P11)
- Limitations:
  - TTS (text-to-speech) was unnatural, making it difficult to focus. (P3, P5, P7, P9, P10, P11)
  - Text-based feedback was suggested as an alternative for better comprehension. (P2, P3, P4, P6, P10, P12)
  - Some experienced information overload when verbal and highlights were combined. (P3, P4, P5, P6, P7, P8, P10, P12)

B.4 Preferred Feedback Modality

- Most users preferred BV (Basic ViF + VeF) or HV (Highlighted ViF + VeF). (P1, P2, P3, P4, P5, P6, P9, P10, P11)
- HV provided the most comprehensive guidance, but some preferred BV to avoid distractions. (P1, P3, P5, P7, P8, P10, P12)
- HN (Highlighted ViF Only) was not sufficient for understanding how to correct mistakes. (P2, P3, P4, P6, P8, P10)
- BN (Basic ViF Only) was the least effective. (P1, P2, P3, P4, P6, P9, P10, P12)

B.5 Suggested Improvements

- Real-time feedback instead of delayed corrections. (P1, P4, P5, P7, P9, P10, P11, P12)
- More natural and expressive TTS voice for better engagement. (P3, P5, P7, P9, P10, P11)
- Improved UI controls:
  - Users should be able to adjust camera angles. (P4, P6, P9, P11, P12)
  - Playback speed control, pause, and rewind functions for VeF. (P1, P3, P5, P7, P10, P11)
- Stronger visual indicators:
  - Numerical values, directional arrows, and enhanced color contrast for highlights. (P2, P4, P5, P7, P10, P12)
  - Potential for haptic feedback to simulate physical corrections. (P3, P5, P7, P9, P12)

B.6 Additional Feedback on System Experience

- AR device limitations:
  - Headset was too heavy, causing discomfort. (P3, P7, P9, P11)
  - Field of view was too narrow, making tracking difficult. (P4, P6, P9, P11)
- Potential integration with golf training facilities:
  - Users suggested applying the system to screen golf simulators for real-world practice. (P5, P7, P9, P12)

B.7 Summary of Key Themes

Received 23 January 2025

Table 3: Verbal Feedback Examples for Each Step

Stage	Explanation
Address	*At the address, your right side moves too far back. This creates an upright posture and reduces stability. Bend slightly at the hips to maintain a forward-leaning spine. Relax your knees to stay grounded, and center your weight evenly over the middle of both feet for better balance.*
Toe-up	*At the toe-up stage, your right elbow moves too far back. Your right shoulder also rotates excessively backward. This pulls the swing path too far inside. An inside path often leads to a flatter swing plane. A flatter plane makes it harder to square the clubface at impact. These issues can cause improper direction of your hit. Keep your right elbow in front of your torso for better connection. Let the club rise naturally along the target line to maintain a better path.*
Backswing	*During the backswing, your upper body shifts too far to the right. This reduces stability and power. Keep your weight on the inside of your right foot. This creates a stable base for better energy transfer. It also helps your body and arms move in sync.*
Top	*At the top of your swing, your left wrist moves backward. This drops the clubface into an open position. An open clubface affects accuracy and consistency. Maintain a flat left wrist to align the clubface properly. A balanced top position improves control and power.*
Downswing	*At the start of the downswing, your left wrist moves backward. Your right shoulder also moves slightly backward. These movements disrupt the natural swing sequence. This misalignment affects the swing path and reduces power. Start the downswing with your hips to guide the sequence. Keep your wrists and shoulders stable as the body rotates.*
Impact	*At impact, your wrists tilt to the right. This opens the clubface and leads to inaccurate shots. A neutral wrist position helps keep the clubface square. Focus on rotating your body through impact. Avoid relying on your wrists for control.*
Full-Motion	*There are two main issues affecting your swing consistency and power. First, your right arm is not fully extended during the backswing. This often occurs when you fail to complete a full shoulder turn when the club goes up. To correct this, focus on achieving a complete shoulder turn, ensuring your shoulders rotate at least 90 degrees at the top stage. Focus on a full arm extension to ensure proper form. Second, your head and wrist are unstable during the swing. This occurs when key points at each stage of the swing are not clearly identified and executed. To address this, concentrate on keeping your head steady by focusing your eyes on the impact point. Additionally, work on maintaining a stable wrist position throughout the swing. Repeated and focused practice will help you build better control and consistency. By addressing these two issues in your practice, you can achieve a more powerful, stable, and consistent swing over time.*

Table 4: Summary of Key Themes in User Interviews

Theme	Key Findings
Most Challenging Feedback Modality	BN was the hardest; HV had mixed reviews due to information overload.
Effectiveness of Highlight Feedback	Helped identify motion differences, but needed numerical and directional improvements.
Effectiveness of Audio Feedback	Provided essential guidance but needed better TTS and text-based alternatives.
Preferred Feedback Modality	BV and HV were most preferred; BN was least effective.
Suggested Improvements	Real-time feedback, better UI controls, stronger visual indicators, and haptic feedback.
System Experience	AR limitations in weight and field of view; potential integration with screen golf application.