Voice to Victory: Modeling In-Game Voice Communication for Team Coordination in MOBA Games

Yongchan Son
Chung-Ang University
Seoul, Republic of Korea
daniel0801@cau.ac.kr

Eunji Park*
Chung-Ang University
Seoul, Republic of Korea
eunjipark@cau.ac.kr

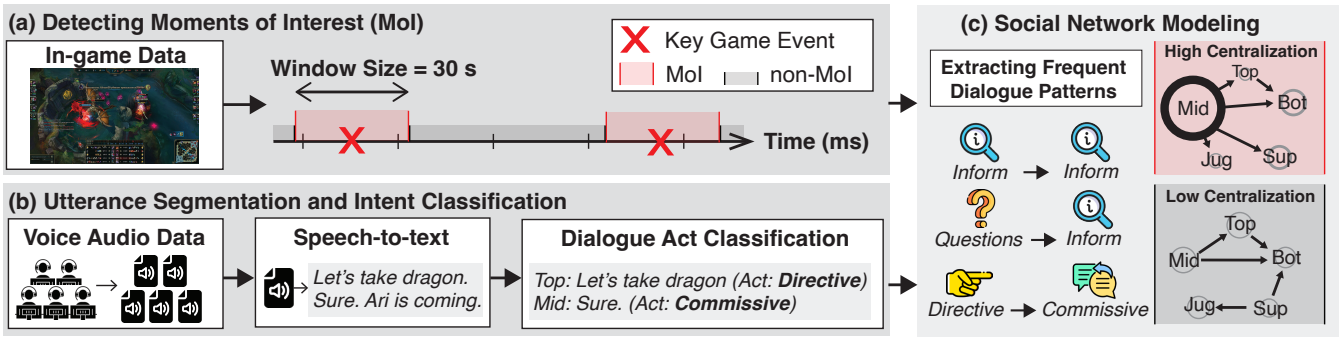

Figure 1: System overview of Voice to Victory

Abstract

In Multiplayer Online Battle Arena (MOBA) games, effective communication is crucial for coordinating team strategies to achieve victory. The influence of communication on team coordination have been examined by prior studies; however, were limited to non-verbal cues. We propose an end-to-end speech act recognition pipeline that automatically infers speakers' intent from voice audio, enabling analysis at scale. Additionally, we applied social network analysis to model verbal communication during team coordination in MOBA. By conducting a data collection study on the gameplay of five players in League of Legends (LoL), we examined team communication was predominantly driven by specific players in collaboratively intensive situations, especially during question-answering interaction.

CCS Concepts

• **Human-centered computing** → *Empirical studies in HCI*.

Keywords

Verbal communication, Coordination, MOBA, Esports

ACM Reference Format:
Yongchan Son and Eunji Park. 2025. Voice to Victory: Modeling In-Game Voice Communication for Team Coordination in MOBA Games. In *The 38th Annual ACM Symposium on User Interface Software and Technology (UIST Adjunct '25), September 28–October 01, 2025, Busan, Republic of Korea.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3746058.3758425

*Corresponding author

UIST Adjunct '25, Busan, Republic of Korea

ACM ISBN 979-8-4007-2036-9/25/09
https://doi.org/10.1145/3746058.3758425

1 Introduction

Research on Multiplayer Online Battle Arena (MOBA) games has gained increasing attention in recent years [2, 5, 19]. Although MOBA games, such as League of Legends (LoL), emphasize team collaboration, existing studies have focused mainly on quantifying individual players' physical and cognitive behaviors [10, 16]. As tactical coordination and communication are crucial for team performance [26], team members continuously exchange task-relevant information, instructions, and emotions via voice chat during gameplay [23]. Although prior work has attempted to demonstrate the importance of communication in Esports, most approaches rely on non-verbal signals (e.g., ping, emote) [12, 25, 27], which lack semantic richness. Rich communication data such as in-game voice chat remains underutilized due to the manual annotation required, limiting its scalability for large-scale analyses.

Motivated by this, we introduce an end-to-end speech act recognition pipeline that automatically processes audio data of players' voices to infer players' communicative intent. Furthermore, we recognized three frequent patterns of dialogue acts based on T-pattern detection [15]. We then construct directed social networks for each pattern, where each player is represented as a node, to quantify distribution of communication among players via the centralization metric [6]. Finally, the contributions of this study can be summarized as follows:

- We developed a framework that automatically classifies speakers' intent from raw voice data.
- We proposed a novel methodology to model verbal communication in Esports environments by applying social network analysis.
- We verified players' participation imbalance in communication varies in highly collaborative situations.

2 Implementation

2.1 Defining Moments of Interest (MoI)

We defined collaboratively intensive tasks as key game events such as kills that involve two or more players, elite monster (e.g., baron, dragon) kill. In such situations, players need to make decisions based on more intensive communication and collaboration. Following prior research [7], we refer to conversations within a 30-second time window (i.e., 15 seconds before and after a key game event) as "Moments of Interest"(MoI) [17], as shown in Figure 1 (a).

2.2 Speech Act Recognition

2.2.1 ***Speech Transcription and Utterance Segmentation.*** Speech transcription pipeline was implemented to produce transcriptions with accurate timestamps in milliseconds. As we collected the players' voice data separately, the following processing procedure was applied individually to all five players' data.

We first transcribed the audio data to utterance using Whisper large-v3 [20], a state-of-the-art ASR model. Given the fast-paced and fragmented nature of in-game voice conversations, obtaining timestamps precisely at utterance level is challenging but crucial for accurate synchronization with gameplay data. To refine the alignment between the transcribed text and the corresponding voice audio, we applied a Connectionist Temporal Classification (CTC) [9] model. To enable its use as input for dialogue act classification, each resulting utterance was translated into English using Gemini[24].

2.2.2 ***Dialogue Act Classification.*** As proposed in He et al. [11], we combined utterance embeddings encoded from a pretrained language model RoBERTa [14] with speaker turn embeddings. These embeddings are then fed to Bi-GRU [3] model to infer Dialogue Acts (DA). DA are higher level semantic abstractions of utterances that conveys the speaker's intent (e.g., Inform, Question, Directive, Commissive) [21]. The model was trained using DailyDialogue (DYDA) dataset [13] for 10 epochs with learning rate of 0.0001, dropout 0.5. Test accuracy of 0.866 was achieved under such conditions. System hierarchy of speech act recognition is shown on Figure 1 (b).

3 Social Network Analysis

3.1 Communication Network Modeling

We computed social networks to quantitatively analyze utterances exchanged between players during conversations [22], as networks are considered a valid tool to investigate complex dynamics in team sports [18]. We defined a social network as a weighted directed graph, where nodes represent players and edges link the player of an utterance to that of the next. Edge weights correspond to the frequency of corresponding utterance. Network was generated from utterance observed during 30000 ms window for each MoI.

3.2 Centralization

We adopted centralization to measure the degree to which a network's structure is dominated by a few highly connected individuals.

	C_{OD}			C_{ID}		
	MoI Mean (SD)	Non-MoI Mean (SD)	p	MoI Mean (SD)	Non-MoI Mean (SD)	p
All	0.31 (0.11)	0.19 (0.02)	0.09	0.32 (0.11)	0.20 (0.02)	0.08
I → I	0.23 (0.04)	0.17 (0.01)	0.12	0.25 (0.04)	0.19 (0.01)	0.09
Q → I	0.26 (0.06)	0.16 (0.004)	**0.01***	0.28 (0.06)	0.19 (0.01)	**0.03***
D → C	0.27 (0.06)	0.15 (0.002)	**0.02***	0.24 (0.06)	0.16 (0.01)	0.33

Table 1: Comparison of centralization between Moments of Interest (MoI) and non-MoI. The 'All' represents all utterances exchanged between players, regardless of dialogue act type. I, Q, D, and C refer to Inform, Question, Directive, and Commissive dialogue acts, respectively. For example, 'Q → I' indicates that a Question utterance was followed by an Inform utterance. (* indicates $p < .05$)

Hence node centralization is defined as follows:

$$S_{OD}(i) = \sum_{j=1}^{N} w_{ij}, \quad C_{OD} = \frac{\sum_{i=1}^{N}(S_{OD}^{\max} - S_{OD}(i))}{(N-1)U}$$

$$S_{ID}(i) = \sum_{j=1}^{N} w_{ji}, \quad C_{ID} = \frac{\sum_{i=1}^{N}(S_{ID}^{\max} - S_{ID}(i))}{(N-1)U}$$

where S_i is node strength [1], defined as sum of weights w_{ij} and $S_{\max}$ represent maximum strength observed across network. N represents the number of players, and U reflects the total number of utterances present in the graph. The denominator serves to standardize the centralization score, mitigating the influence of the total number of utterances on the metric. Outdegree centralization C_{OD} captures concentration of communication initiation (e.g., giving directives) of communicators, while indegree centralization C_{ID} reflects how centralized responses (e.g., replying to directives) are by receivers. See Figure 1 (c) for examples of social networks analysis.

4 Data Acquisition

A total of five male participants were recruited, all having extensive experience playing LoL, with an average of 11 years (SD=1.58). Participants engaged in six flex queue matches where players communicated freely via Discord. All voice interactions were recorded using the Craig bot [4]. In-game data (e.g., key event log, kill/assist record) was logged via Live Client API [8], an API that enables collecting data directly from the League of Legends client.

5 Result

As shown in Table 1, both mean of C_{ID} and C_{OD} were consistently higher during MoI. Specifically, indegree and outdegree centralization for Question → Inform pattern were significantly higher during MoI. These findings suggest that under circumstances requiring rapid information flow and cooperation, specific individuals tended to assume dominant roles in both seeking and providing information, thereby enhancing communicative efficiency.

In addition, outdegree centralization in Directive → Commissive exchanges was significantly higher during MoI. One possible explanation is that commands were mainly issued by a limited set of individuals, which might have contributed to minimizing confusion and maintaining strategic consistency in high-pressure situations.

6 Conclusion

In this study, we proposed an automatic pipeline that infers speakers' intent from voice audio. We explored the distribution of communication between players by evaluating centralization from social network. Additionally, we assessed centralization difference during highly collaborative situations, referred to as MoI. By analyzing the collected data, we confirmed communication involving question and answering was more centralized during MoI. In future work, we are planning to collect additional data from Esports professional gamers for longitudinal analysis.

References

[1] Alain Barrat, Marc Barthelemy, Romualdo Pastor-Satorras, and Alessandro Vespignani. 2004. The architecture of complex weighted networks. *Proceedings of the national academy of sciences* 101, 11 (2004), 3747–3752.

[2] Tom Batsford. 2014. Calculating optimal jungling routes in dota2 using neural networks and genetic algorithms. *Game Behaviour* 1, 1 (2014).

[3] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259* (2014).

[4] Coolguy3289 and Snazzah. 2025. https://craig.chat/

[5] Anders Drachen, Matthew Yancey, John Maguire, Derrek Chu, Iris Yuhui Wang, Tobias Mahlmann, Matthias Schubert, and Diego Klabajan. 2014. Skill-based differences in spatio-temporal team behaviour in defence of the ancients 2 (dota 2). In *2014 IEEE Games Media Entertainment*. IEEE, 1–8.

[6] Linton C Freeman et al. 2002. Centrality in social networks: Conceptual clarification. *Social network: critical concepts in sociology. Londres: Routledge* 1, 3 (2002), 238–263.

[7] Julian Frommel and Regan L Mandryk. 2024. Toxicity in online games: The prevalence and efficacy of coping strategies. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–12.

[8] Riot Games. 2025. Riot Developer Portal. https://developer.riotgames.com/docs/lol.

[9] Alex Graves. 2012. Connectionist temporal classification. In *Supervised sequence labelling with recurrent neural networks*. Springer, 61–93.

[10] Alyona Grushko, Olga Morozova, Mikhail Ostapchuk, and Ekaterina Korobeynikova. 2021. Perceptual-cognitive demands of esports and team sports: A comparative study. In *Advances in Cognitive Research, Artificial Intelligence and Neuroinformatics: Proceedings of the 9th International Conference on Cognitive Sciences, Intercognsci-2020, October 10-16, 2020, Moscow, Russia 9*. Springer, 36–43.

[11] Zihao He, Leili Tavabi, Kristina Lerman, and Mohammad Soleymani. 2021. Speaker turn modeling for dialogue act classification. *arXiv preprint arXiv:2109.05056* (2021).

[12] Alex Leavitt, Brian C Keegan, and Joshua Clark. 2016. Ping to win? Non-verbal communication and team performance in competitive online multiplayer games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 4337–4350.

[13] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. Dailydialog: A manually labelled multi-turn dialogue dataset. *arXiv preprint arXiv:1710.03957* (2017).

[14] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).

[15] Magnus S Magnusson. 2000. Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior research methods, instruments, & computers* 32, 1 (2000), 93–110.

[16] Eugen Nagorsky and Josef Wiemeyer. 2020. The structure of performance and training in esports. *PloS one* 15, 8 (2020), e0237584.

[17] Amin Noroozi, Mohammad S Hasan, Maryam Ravan, Elham Norouzi, and Ying-Ying Law. 2024. An efficient machine learning approach for extracting eSports players' distinguishing features and classifying their skill levels using symbolic transfer entropy and consensus nested cross-validation. *International Journal of Data Science and Analytics* (2024), 1–14.

[18] Pedro Passos, Keith Davids, Duarte Araújo, N Paz, J Minguéns, and Jose Mendes. 2011. Networks as a novel tool for studying team ball sports as complex social systems. *Journal of science and medicine in sport* 14, 2 (2011), 170–176.

[19] Nataliia Pobiedina, Julia Neidhardt, Maria del Carmen Calatrava Moreno, Laszlo Grad-Gyenge, and Hannes Werthner. 2013. On successful team formation: Statistical analysis of a multiplayer online game. In *2013 IEEE 15th Conference on Business Informatics*. IEEE, 55–62.

[20] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *International conference on machine learning*. PMLR, 28492–28518.

[21] John R Searle. 1969. Speech acts: An essay in the philosophy of language. *Cambridge University* (1969).

[22] Shazia Tabassum, Fabiola SF Pereira, Sofia Fernandes, and João Gama. 2018. Social network analysis: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8, 5 (2018), e1256.

[23] Evelyn TS Tan, Katja Rogers, Lennart E Nacke, Anders Drachen, and Alex Wade. 2022. Communication sequences indicate team cohesion: A mixed-methods study of Ad Hoc league of legends teams. *Proceedings of the ACM on Human-Computer Interaction* 6, CHI PLAY (2022), 1–27.

[24] Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805* (2023).

[25] Jason Wuertz, Scott Bateman, and Anthony Tang. 2017. Why players use pings and annotations in Dota 2. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 1978–2018.

[26] Pu Yang, Brent E Harrison, and David L Roberts. 2014. Identifying patterns in combat that are predictive of success in MOBA games.. In *FDG*.

[27] Keyang Zheng, Ben Stein, and Rosta Farzan. 2023. Use ping wisely: A study of team communication and performance under lean affordance. *ACM Transactions on Social Computing* 5, 1-4 (2023), 1–26.