Augmenting Online Meetings with Context-Aware Real-time Music Generation

Haruki Suzawa
harukisuzawa@gmail.com
Nara Institute of Science and Technology
Nara, Japan

Ko Watanabe
ko.watanabe@dfki.de
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, Germany

Andreas Dengel
andreas.dengel@dfki.de
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, Germany

Shoya Ishimaru
ishimaru@omu.ac.jp
Osaka Metropolitan University
Osaka, Japan

Abstract

As online communication continues to expand, participants often face cognitive fatigue and reduced engagement. Cognitive augmentation, which leverages technology to enhance human abilities, offers promising solutions to these challenges. In this study, we investigate the potential of generative artificial intelligence (GenAI) for real-time music generation to enrich online meetings. We introduce *Discussion Jockey 2*, a system that dynamically produces background music in response to live conversation transcripts. Through a user study involving 14 participants in an online interview setting, we examine the system's impact on relaxation, concentration, and overall user experience. The findings reveal that AI-generated background music significantly enhances user relaxation (average score: 5.75/9) and concentration (average score: 5.86/9). This research underscores the promise of context-aware music generation in improving the quality of online communication and points to future directions for optimizing its implementation across various virtual environments.

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**; *Collaborative and social computing*; Interaction paradigms; User studies.

Keywords

Generative AI, Music Generation, Online Meeting, Communication

ACM Reference Format:
Haruki Suzawa, Ko Watanabe, Andreas Dengel, and Shoya Ishimaru. 2024. Augmenting Online Meetings with Context-Aware Real-time Music Generation. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym '24)*. ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

Conference acronym '24, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/2024/06
https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Cognitive augmentation leverages technology to enhance human cognitive abilities [1, 13]. This concept is particularly applied in mobile computing devices [5, 13, 19, 20]. Our society now operates in a hybrid mode, where offline and online communication blend physical and virtual interactions. This transition is driven by technological advancements and the growing need for flexible ways to connect and collaborate. In this new era, cognitive augmentation holds significant potential for enhancing communication. For example, in online communication, cognitive augmentation tools are used to support memory retention [10], boost engagement [16, 17], facilitate smooth interactions [8, 15], and assist in decision-making processes [7]. These tools range from simple note-taking applications [14] that enhance memory retention to advanced AI-driven systems providing real-time feedback and suggestions [12], transforming how we conduct and participate in meetings.

Recent generative artificial intelligence (GenAI) advancements offer new possibilities for augmenting meetings [11, 12]. For example, Son et al. [14] investigated real-time transcript summarization during meetings using large language models (LLMs) like BERT [6]. The goal was to see if such summarization could help participants focus on the meeting content. The results show that real-time transcription summarization with LLMs helps participants maintain focus. Another GenAI application is presented by Rajaram et al. [12], who developed a system called *BlendScape* that uses GenAI to personalize video-conferencing environments. This application collects transcripts during meetings and uses them as prompts to generate background images, creating a customized video-conferencing environment for an immersive experience. Among these approaches, real-time music generation remains unexplored in the context of augmenting meetings.

Music profoundly impacts human emotions and can be a powerful tool for healing and relaxation [2, 3]. The right music can create a calming atmosphere, reduce anxiety, and improve overall mood [9]. Background music can also help improve focus and concentration [18]. Concerning the previous findings, our study aims to use music GenAI to consider the context for creating a personalized and practical auditory experience—the strategic use of music in online interviews and discovering how beneficial for humans.

In this study, we focus on evaluating the effectiveness of GenAI for music generation in online meetings. Our target scenario is an

Table 1: Comparison of related work for augmenting online meetings. The "Context Aware" column represents a checklist indicating whether the research utilizes the meeting transcript. The "GenAI" column represents a checklist indicating whether the research use GenAI technology. The "Music" column represents a checklist indicating whether the research focuses on using music in meetings.

Author	Context Aware	GenAI	Music	Performance Detail
Son et al. [14]	✓	✓	✗	Real-time summarization of the meeting transcripts using LLMs (BERT).
Park et al. [11]	✓	✓	✗	Introduce *CoExplorer2D* and *CoExplorerVR* managing meeting progress.
Han et al. [4]	✓	✓	✗	Collaborative creation of concept-based image generation (Midjourney).
Rajaram et al. [12]	✓	✓	✗	Real-time generation of a meeting background images.
Feng et al. [3]	✗	✗	✓	Collaborative music creation for therapy using the prototype *ComString*.
Suzawa et al. [15]	✗	✗	✓	A different metronome (BPM) is sounded in real-time for each participant.
Ours	✓	✓	✓	Real-time generation of a personalized music using transcript as a prompt.

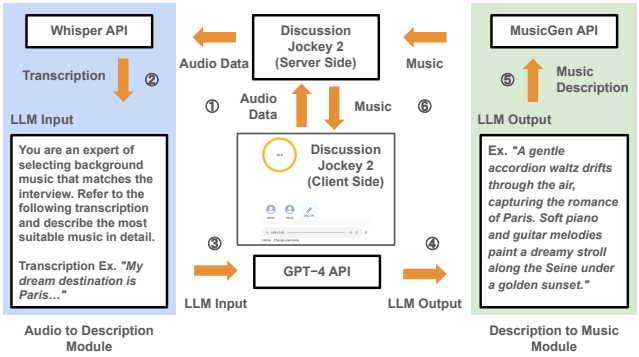

Figure 1: Proposed architecture of *Discussion Jockey 2*. The application uses Whisper API to collect speech transcripts. The transcript and template prompt are input in GPT-4 to generate a music description prompt optimized for MusicGen API. The generated music is then feedback to the application and played in the participant interface.

online interview setting where the interviewer and interviewee are in different locations. By collecting transcripts during the online interview, we generate music that considers the context of the interview and personalizes the music. Our demonstration aims to explore the following research questions: (RQ1) Does context-aware music generation make participants feel relaxed in online meetings? (RQ2) Does context-aware music generation enhance participants' concentration in online meetings?

2 Related Work

Table 1 highlights the uniqueness of our research in comparison to existing studies. Several researchers have explored context-aware meeting augmentation [4, 11, 12, 14]. For instance, Son et al. [14] and Park et al. [11] concentrate on summarizing or facilitating meetings through transcripts, utilizing LLMs to generate summaries or assist in meeting progression. Conversely, Han et al. [4] and Rajaram et al. [12] employ image generation to enhance meetings. Specifically, Han et al. [4] use GenAI for real-time collaborative image creation during discussions, while Rajaram et al. [12] generate background images based on speech context to enhance immersion. Although these studies utilize GenAI, none focus on context-aware music generation for online meeting augmentation.

In the realm of music used during online meetings, Feng et al. [3] propose the collaborative creation of therapeutic music using *ComString*, which involves Collaborative Digital Musical Instruments (CDMIs) for real-time music composition. Their work aims to generate music for therapy, not general meeting contexts, and does not incorporate transcripts. Our research closely relates to Suzawa et al. [15], where metronome (BPM) music is produced based on participants' speech data, with higher BPM for more active speakers and lower for quieter ones. This study aims to balance speaking time among participants but does not focus on the context or content of the transcripts. Both research uses music as a target in the meeting, but neither work uses transcripts nor GenAI.

In conclusion, while various researchers have explored context-aware online meeting augmentation and the use of music in meetings, the context-aware generation of music for meeting enhancement remains underexplored. Therefore, our research aims to investigate the impact of applying context-aware music generation in online meetings.

3 Methodology

Figure 1 illustrates the architecture of *Discussion Jockey 2*, a system that generates real-time background music based on live conversation. The client-side, built with React and WebSocket, facilitates seamless user-server interaction. The workflow of the server-side consists of three stages: (1) transcribing real-time audio using Whisper API [1], (2) generating a music description via GPT-4 after accumulating three minutes of transcription, and (3) creating a customized music track with MusicGen API [2]. The description defines tempo, style, and instrumentation, ensuring dynamic adaptation to conversation context. To reduce music generation time, the music length was set to 10 seconds and looped for 3 minutes. For instance, when the transcription includes *"My dream destination is Paris..."*, the system generates a fitting music description, as shown in Figure 1: *"A gentle accordion waltz drifts through the air, capturing the romance of Paris. Soft piano and guitar melodies paint a dreamy stroll along the Seine under a golden sunset."*.

We conducted a 20-minute one-on-one interview experiment with 14 participants, as shown in Figure 2. Only the interviewee's voice data was processed. Background music was introduced six minutes in and changed every three minutes, allowing participants

[1] https://openai.com/research/whisper
[2] https://replicate.com/meta/musicgen/api

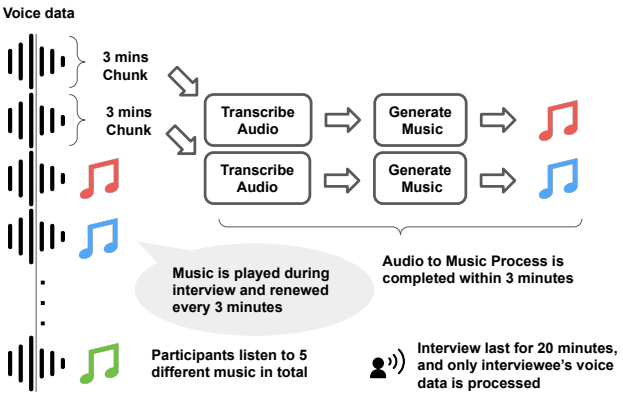

Figure 2: Experiment Design

to experience five different pieces. To maintain a natural and engaging conversation, they were asked casual questions about hobbies, travel, and favorite foods. After listening to all five pieces, they evaluated each based on three criteria: whether it helped them relax, improved concentration, and whether they liked it.

4 Result & Discussion

To explore the research questions (RQs) and identify the most beneficial scenarios for the proposed system, we analyzed subjective evaluations of the generated music. The average ratings were 5.75 for relaxation (1: very nervous, 9: very relaxed) and 5.86 for concentration (1: very distracted, 9: very concentrated), indicating a tendency to enhance both. In this experiment, the system generated music that matched the conversation content (Figure 1, LLM Input). However, modifying the prompt can generate music tailored to enhance relaxation or concentration.

The average response to "Did you like the music?" (1: hated, 9: liked) was 5.67, indicating a generally positive reception. A follow-up survey collected 15 responses, with eight citing musical preferences (tune, tempo, melody, mood) as key factors. Participants also wanted their preferences for personalized music to be considered in advance. The next prototype will incorporate genre, nationality, gender, and age to assess the impact on concentration and relaxation during meetings. Three participants prioritized relaxation, while two emphasized the importance of situational and emotional alignment. Currently, the system generates music based on transcriptions from two chunks earlier (each three minutes long) to allow for processing time. However, enhancing real-time performance will be a critical area for future improvement. Additionally, two participants prioritized concentration, underscoring the need for non-disruptive music. As for volume, responses to "How loud was the sound played?" (1: too quiet, 10: too loud) averaged 6.60, indicating that the volume was relatively high and may have influenced responses. Meanwhile, "Did you find the frequency of music transitions comfortable?" (1: very uncomfortable, 10: very comfortable) received an average score of 6.65, suggesting that the three-minute update interval was generally well-received.

To understand suitable usage scenarios for this system, participants were asked: "In what situations would you like to use this system?". A total of 13 responses were collected. The most common category (six responses) involved individual tasks like working, studying, driving, and meditating, with an emphasis on using the system for boring tasks. Reading was specifically mentioned, suggesting future research on generating music based on book content. Four responses favored using the system in meetings, with two highlighting its role in setting the mood for casual conversations with friends. Two respondents offered potential educational uses: one suggested using it for relaxation during presentation, and the other advocated for maintaining focus during lectures—both underscoring its value in an educational context.

5 Conclusion

This study introduced *Discussion Jockey 2*, a novel approach to augmenting online meetings through real-time, context-aware music generation. By leveraging GenAI technologies, our system personalizes background music based on meeting transcripts, aiming to improve relaxation and concentration during virtual interactions. Our findings suggest that AI-generated music can create a more engaging and comfortable environment, with most participants reporting positive effects. However, personalization and real-time processing remain key factors in optimizing user experience, as individual preferences significantly influence perception. Future work will focus on refining music customization based on user-specific attributes, such as nationality, age, and music preferences, to enhance the adaptability and impact of AI-driven auditory augmentation in virtual meetings.

Acknowledgments

This work is supported by the OMU Project "Verifying the Health-Promoting Effects of Music as a Social Prescription".

References

[1] Sarah Clinch and Jamie A Ward. 2023. Augmented Cognition. *IEEE Pervasive Computing* 22, 3 (2023), 6–7.
[2] Claudius Conrad. 2010. Music for healing: from magic to medicine. *The Lancet* 376, 9757 (2010), 1980–1981.
[3] Yuan-Ling Feng, Zhaoguo Wang, Yuan Yao, Hanxuan Li, Yuting Diao, Yu Peng, and Haipeng Mi. 2024. Co-designing the Collaborative Digital Musical Instruments for Group Music Therapy. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 698, 18 pages. doi:10.1145/3613904.3642649
[4] Yuanning Han, Ziyi Qiu, Jiale Cheng, and RAY LC. 2024. When Teams Embrace AI: Human Collaboration Strategies in Generative Prompting in a Creative Design Task. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 176, 14 pages. doi:10.1145/3613904.3642133
[5] Shoya Ishimaru, Nicolas Großmann, Andreas Dengel, Ko Watanabe, Yutaka Arakawa, Carina Heisel, Pascal Klein, and Jochen Kuhn. 2018. HyperMind Builder: Pervasive User Interface to Create Intelligent Interactive Documents. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (Singapore, Singapore) *(UbiComp '18)*. Association for Computing Machinery, New York, NY, USA, 357–360. doi:10.1145/3267305.3267667
[6] Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT*, Vol. 1. Minneapolis, Minnesota.
[7] Hanseob Kim, Bin Han, Jieun Kim, Muhammad Firdaus Syawaludin Lubis, Gerard Jounghyun Kim, and Jae-In Hwang. 2024. Engaged and Affective Virtual Agents: Their Impact on Social Presence, Trustworthiness, and Decision-Making in the Group Discussion. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 149, 17 pages. doi:10.1145/3613904.3642917

[8] Mikko Kytö, Ilyena Hirskyj-Douglas, and David McGookin. 2021. From Strangers to Friends: Augmenting Face-to-face Interactions with Faceted Digital Self-Presentations. In *Proceedings of the Augmented Humans International Conference 2021* (Rovaniemi, Finland) *(AHs '21)*. Association for Computing Machinery, New York, NY, USA, 192–203. doi:10.1145/3458709.3458954

[9] Rollin McCraty, Bob Barrios-Choplin, Michael Atkinson, and Dana Tomasino. 1998. The effects of different types of music on mood, tension, and mental clarity. *Alternative therapies in health and medicine* 4, 1 (1998), 75–84.

[10] Takato Mizuho, Tomohiro Amemiya, Takuji Narumi, and Hideaki Kuzuoka. 2023. Virtual Omnibus Lecture: Investigating the Effects of Varying Lecturer Avatars as Environmental Context on Audience Memory. In *Proceedings of the Augmented Humans International Conference 2023* (Glasgow, United Kingdom) *(AHs '23)*. Association for Computing Machinery, New York, NY, USA, 55–65. doi:10.1145/3582700.3582709

[11] Gun Woo Park, Payod Panda, Lev Tankelevitch, and Sean Rintel. 2024. The CoExplorer Technology Probe: A Generative AI-Powered Adaptive Interface to Support Intentionality in Planning and Running Video Meetings. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 1638–1657.

[12] Shwetha Rajaram, Nels Numan, Balasaravanan Thoravi Kumaravel, Nicolai Marquardt, and Andrew D Wilson. 2024. BlendScape: Enabling End-User Customization of Video-Conferencing Environments through Generative AI. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 40, 19 pages. doi:10.1145/3654777.3676326

[13] Albrecht Schmidt. 2017. Augmenting Human Intellect and Amplifying Perception and Cognition. *IEEE Pervasive Computing* 16, 1 (2017), 6–10. doi:10.1109/MPRV.2017.8

[14] Seoyun Son, Junyoug Choi, Sunjae Lee, Jean Y Song, and Insik Shin. 2023. It is Okay to be Distracted: How Real-time Transcriptions Facilitate Online Meeting with Distraction. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 64, 19 pages. doi:10.1145/3544548.3580742

[15] Haruki Suzawa, Ko Watanabe, Masakazu Iwamura, Koichi Kise, Andreas Dengel, and Shoya Ishimaru. 2023. Supporting Smooth Interruption in a Video Conference by Dynamically Changing Background Music Depending on the Amount of Utterance. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers* (Cambridge, United Kingdom) *(UbiComp/ISWC '22 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 299–302. doi:10.1145/3544793.3560384

[16] Ko Watanabe, Andreas Dengel, and Shoya Ishimaru. 2024. MetacognitionEnGauge: Real-time Augmentation of Self-and-Group Engagement Levels Understanding by Gauge Interface in Online Meetings. In *Proceedings of the Augmented Humans International Conference 2024* (Melbourne, VIC, Australia) *(AHs '24)*. Association for Computing Machinery, New York, NY, USA, 301–303. doi:10.1145/3652920.3653054

[17] Ko Watanabe, Tanuja Sathyanarayana, Andreas Dengel, and Shoya Ishimaru. 2023. EnGauge: Engagement Gauge of Meeting Participants Estimated by Facial Expression and Deep Neural Network. *IEEE Access* 11 (2023), 52886–52898. doi:10.1109/ACCESS.2023.3279428

[18] Kevin JP Woods, Gonçalo Sampaio, Tedra James, Emily Przysinda, Adam Hewett, Andrea E Spencer, Benjamin Morillon, and Psyche Loui. 2024. Rapid modulation in music supports attention in listeners with attentional difficulties. *Communications Biology* 7, 1 (2024), 1376.

[19] Kanta Yamaoka, Ko Watanabe, Koichi Kise, Andreas Dengel, and Shoya Ishimaru. 2023. Experience is the Best Teacher: Personalized Vocabulary Building Within the Context of Instagram Posts and Sentences from GPT-3. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers* (Cambridge, United Kingdom) *(UbiComp/ISWC '22 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 313–316. doi:10.1145/3544793.3560382

[20] Kanta Yamaoka, Ko Watanabe, Koichi Kise, Andreas Dengel, and Shoya Ishimaru. 2025. Img2Vocab: Explore Words Tied to Your Life With LLMs and Social Media Images. *IEEE Access* 13 (2025), 20456–20471. doi:10.1109/ACCESS.2025.3533076

Received 20 February 2024; revised 12 March 2024; accepted 5 June 2024