Designing the Interactive transition of Abstraction and Concretization in Creative Ideation: A Case Study on Lyric Composition

ANONYMOUS AUTHOR(S)

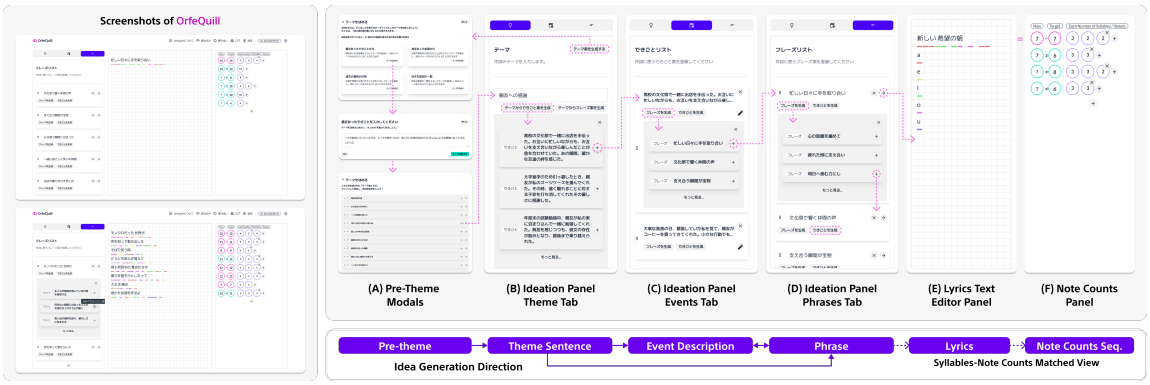

Fig. 1. Application screenshots and descriptions of the UI panels correspond to the flow of lyrics writing in OrfeQuill. Each type of idea can be generated in several directions and is mounted on each tab of the **Ideation Panel** that pools them as cards.

This paper presents OrfeQuill, a novel lyrics writing support system that leverages Large Language Models (LLMs) to assist users in various stages of the lyrics writing process. While existing tools often generate lyrics from a single theme or prompt, such one-step approaches overlook the deeper layers of thought and emotion inherent in the creative process and may cause idea fixation. Therefore, OrfeQuill, adopts a multi-step approach by providing users with three stages of ideation with an LLM: setting a theme derived from personal experiences, identifying key story events, and finally composing lyric phrases. This incremental workflow encourages lyricists to explore wider ideas and reflect them more carefully on expression, enabling them to align the final output with their authentic emotions and intentions. To validate our approach, we conducted a preliminary evaluation with six experienced lyricists. The results suggest that OrfeQuill not only fosters creative exploration but also helps streamline the writing process, offering valuable insights into how AI-assisted multi-step support can enhance the songwriting process.

CCS Concepts: • **Human-centered computing** → **User centered design**; • **Applied computing** → *Sound and music computing*.

Additional Key Words and Phrases: Creativity support tools, Writing assistants, Lyrics writing support

1 INTRODUCTION

The rapid advancement of large language models (LLMs) has fueled the development of Creative Support Tools (CSTs)[5], which utilize LLMs to aid users in generating written content across various domains, from narrative composition to academic writing [10]. While studies have shown CSTs to be effective in boosting creativity and streamlining workflows, challenges remain regarding idea fixation[16] and output homogenization[3]. Lyric-writing, a particularly complex form of creative writing, requires not only harmonization with melody, rhythm, and phonetics, but also the conveyance of narrative and deep personal emotions[14, 18, 17]. Despite these complexities, AI-supported lyrics writing interactions remain relatively unexplored, making these interactions the focus of this study.

Current LLM lyric-writing tools instantly generate text resembling song lyrics from simple user inputs, such as themes or moods used as prompts or parameters[19, 11, 9, 13, 1, 2, 20]. This one-step generation requires complex elements to be processed in a complex state, overloading users. Users are forced to frequently regenerate lyrics, review them, and

edit for details, possibly sticking to the initial LLM output. Moreover, users themselves may not fully understand or be able to articulate the complexity of their feelings, making it difficult for one-step generation tools to capture the depth of these emotions accurately only by prompting. Therefore, we propose assisting lyrics writers by supporting the cognitive process of translating thoughts into lyrics. Our multi-perspective editing approach, detailed in Section 2, facilitates broader idea exploration and more accurately reflects the user's intentions, aiming to prevent fixation and homogenization in AI-assisted lyric writing.

In this research, we developed OrfeQuill, a multi-step lyric support system designed to decompose the complex information within lyrics into individually manageable components. It incorporates three core features: 1) Theme Suggestion: Proposing some core themes based on the user's personal experiences and favorite words (referred to as a Pre-theme). 2) Event Suggestion: Assisting in identifying key narrative events related to the chosen theme or event. 3) Lyric Phrase Suggestion: Generating candidate lyric phrases for each event or phrase. This incremental approach allows users to explore and refine ideas, progressively translating them into lyrics that more accurately reflect their personal narratives.

2 LLM-BASED LYRICS WRITING SUPPORT

2.1 Design Overview

Lyrics writing involves multiple, flexible stages of abstraction, from conceptualizing a theme or message to crafting singable phrases. Designed for songwriting workflows that prioritize melody creation, one of the common practice in music production, OrfeQuill leverages an LLM to support the various stages of lyric writing. It consists of three panels (Figure 1): ideation tabs spanning abstraction levels (Figure 1 B-D), a lyric editor (Figure1 E), and a verifying line-to-melody note correspondence (Figure1 F). We implemented it as a web application and utilized GPT-4o API (version 2024-11-20) as a backend.

2.2 Support for ideation at different levels of abstraction

Lyrics convey a worldview or message of the writer, termed **Theme** in OrfeQuill. These are represented as user-defined single sentences (Figure 1 B). To aid theme creation, this application leverages an LLM to generate themes based on user-provided **Pre-themes**, which can be quotes, personal memories, or remarks by others.

Following Theme definition, the narrative and emotional progression are structured through **Events**. These Events, managed in a card user interface, can also be generated by the LLM using the Theme as input (Figure 1 B-C).

Once Events are established, they are translated into concrete **Phrases**. OrfeQuill provides a card user interface for organizing these ideas (Figure 1 D). The LLM can convert Themes or Events into lyrical Phrases, which are then added to a phrase pool (Figure 1 C). Furthermore, existing Phrases can be used to generate new Phrases for rephrasing, and new Events for further storyline development.

2.3 Putting ideas into lyrics

Following the ideation process, it is necessary to compose these phrases to the corresponding lines of the melody. At this time, it is crucial to consider the consistency with the melody, rhyme, and related elements. In this application, users can pre-register melody note counts, which provides an interface for corresponding lyric creation.

A central panel incorporates a block editor that allows for per-line editing. Inspired by several existing editors[12, 8], we incorporated a UI that indicates sequences of mora. Beneath the lyrics, colored bars representing mora counts,

corresponding to vowels, are displayed (Figure 1 E). Also, total number of notes per line is important, as seen in some existing editors[19, 15]. The rightmost panel displays written lyric syllable counts: total syllables of current line, per-line assignments, and per-word assignments. When the number of syllables in a line matches the pre-registered note count, the UI color changes to red (Figure 1 F).

3 PRELIMINARY EVALUATION

In the preliminary evaluation, we confirmed the overall usability of the tool and the effectiveness of the generation function for each phase, and the appropriateness of the abstraction classifications for the lyricists' process. The tool was implemented in Japanese. Six Japanese subjects (two in their 20s, four in their 30s. Five male, one female), all of whom had written lyrics in the past as a hobby (two have 1 to 3 years, and four have at least 3 years of experience in music creation), participated in the study. The test session had 60 minutes, during which participants received 5 minutes of instruction and a 10-minute tutorial, and then a 30-minute free lyrics writing task for a melody. Participants then had 15 minutes to answer a questionnaire, and participate in a semi-structured interview based on the interview guide and their questionnaire answers. Participants selected one of the three types of Pops songs for which they would write lyrics, and the Note Counts Panel (Figure 1 F) was started with the lyrics already filled.

A quantitative evaluation was conducted to identify relative differences among the evaluation metrics and to understand the reasons for these differences. The 5-point Likert scale metrics was sat based on Creativity Support Index[4, 6] and NASA-TLX[7] in order to gain insight into the creative support properties of the current tool, and the lyrics writing task itself. Metrics were *Ease of work*, *Ease of use of the tool*, *Usefulness*, *Productivity*, *Creativity*, *Enjoyment*, *Satisfaction with the results*, *Sense of ownership*, and usefulness of each of the generation functions of Figure 1.

The following qualitative survey interview was based on two questions: *What are your thoughts on the lyric writing experience using OrfeQuill?*, *How did it compare to your usual lyrics writing experience? Is the preference process similar?*. Then these were followed by the reasons for each of the quantitative evaluation items.

4 RESULTS

In the quantitative results, overall average scores were above 3.33, and no major problems were found. Relatively high scores were obtained for *Ease of usage* and *Enjoyment*, while the *Creativity* and *Sense of ownership* of the work were relatively low. The overall results were: *Ease of usage* (4.33 ± 0.52), *Enjoyment* (4.17 ± 0.75), *Helpfulness* (3.83 ± 0.75), *Productivity* (3.83 ± 1.17), *Ease of lyrics task* (3.5 ± 1.22), *Satisfied with the result* (3.5 ± 0.84), *Creativity* (3.33 ± 1.21), and *Sense of ownership* (3.33 ± 1.37).

And for the generation function, the phrase generation function from highly abstract information was well evaluated. On the other hand, phrase-to-phrase generation and event-to-event generation were rated low. The results were: Event-to-Phrase (4.3 ± 0.82), Theme-to-Phrase (4.2 ± 0.96), Phrase-to-Event: (4.0 only one response), Pretheme-to-Theme (3.6 ± 0.82), Theme-to-Event (3.6 ± 0.82), Phrase-to-Phrase (3.5 ± 0.71), Event-to-Event: (2.0, only one response).

From the interviews, we found that the tool covers the various ways of developing conceptual ideas, which differ from user to user, and found requests for new features and current issues were identified.

OrfeQuill's approach of moving between different levels of abstraction in the lyrics seemed natural to many users in terms of thinking about the theme ("*It was a natural process*" (P1), "*I do it like that when I do it myself. I do similar things. The flow of raising the level of abstraction is the same when I do it myself*" (P3), "*It was close to how I feel when I want to write*" (P4), By contrast, several users mentioned that events, in some cases, do not form a concrete event ("*It depends on the person whether the lyrics are in a story-like style or not. For some, the ideological part is the core, for*

others, perhaps a perspective that is not an event is necessary. In chorus sections, for example, I might want to sing about personal beliefs and convictions related to a particular theme." (P5), "*Some of the lyrics are not sung about specific events. The middle part is just one of many ways of making things happen.*" (P6)).

Users found the LLM produced generic phrases, hindering its ability to contribute to innovative lyrics writing. A key limitation identified by participants P4, P5 and P6 is the perceived lack of originality and diversity in the LLM-generated content, which caused relatively low score of *Creativity* and *Sense of Ownership*.

5 DISCUSSION AND FUTURE WORK

We confirmed that the method of gradually building lyrics from a theme into phrases, rather than generating complete lyrics at once, successfully expresses the thoughts during lyrics writing and facilitates the rapid exploration of numerous ideas without thinking about the prompts.

Notably, some users explicitly highlighted the value of the Event ideation feature, which is the unique feature of our tool. For instance, P1 praised the tool's ability to generate an Event from a Phrase, which shifts the abstraction level upward. P1 explained, "*When I get stuck, I can return to a higher-level perspective to refresh my viewpoint.*" This suggests that providing mechanisms to toggle between different levels of abstraction could play a significant role in alleviating creative fixation and reducing the writer's tendency to stick to previously generated content.

On the other hand, we identified several issues and possibilities for future development. The low scores for *Creativity* and *Sense of Ownership* were likely due to the low quality outputs resulting from the simple prompts. However, this was also due to the lack of user control over the generation process. To prevent fixation and homogenization, OrfeQuill needs not only to allow users to quickly explore different levels of abstraction but also to provide the flexibility to utilize a wide range of AI outputs within the given context. In the future, we intend to design this range based on the variables the lyricist wishes to explore, rather than providing a simple text field for prompts. For instance, it would be effective to offer features such as keyword inputs, adjustable ambiguity of expression of each generation. It might also be effective to generate not only from adjacent steps (e.g., Theme-to-Event) but also from other contextual factors, including user-written lyrics, the initial theme and pre-theme, and concept transitions observed in edit logs. Through further user testing and analysis, we plan to better understand the needs of lyricists and their detailed workflow.

Also, intermediate processes should not be represented solely by Events; a more generalized approach encompassing emotions, landscapes, and scenes is necessary. While the current Event representation aims to support lyric timeline construction, incorporating abstract concepts and beliefs alongside concrete events may introduce undue complexity. Interactions beyond simply manipulating the current card-based UI for idea arrangement and result organization are required. We plan to conduct a more in-depth analysis of how the timeline of lyrics is being shaped and the more appropriate interface for externalizing that idea of the timeline in the lyricist's mind.

From a broader perspective, we are interested in idea generation that accommodates the diverse workflows of individual lyricists. To this end, it will be essential to create an environment in which various generative support functions can be dynamically selected and provided by adaptively selecting the context to be considered by the model or by understanding the user's intentions without explicit prompting.

References

[1] Junia AI. 2025. Junia lyric generator. Accessed on 2025-01-20. (2025). https://www.junia.ai/templates/lyric.

[2] Alive Co.,Ltd. 2025. Ai sakkyoku (composition). ja. Accessed on 2025-01-20. (2025). https://www.aisakkyoku.com/.

[3] Barrett R Anderson, Jash Hemant Shah, and Max Kreminski. 2024. Homogenization effects of large language models on human creative ideation. In *Creativity and Cognition*. C&C '24: Creativity and Cognition (Chicago IL USA). ACM, New York, NY, USA, (June 23, 2024). doi:10.1145/3635636.3656204.

[4] Erin A Carroll and Celine Latulipe. 2009. The creativity support index. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '09) (Boston, MA, USA). Association for Computing Machinery, New York, NY, USA, (Apr. 4, 2009), 4009–4014. ISBN: 9781605582474. doi:10.1145/1520340.1520609.

[5] Tuhin Chakrabarty, Vishakh Padmakumar, Faeze Brahman, and Smaranda Muresan. 2024. Creativity support in the age of large language models: An empirical study involving professional writers. In *Creativity and Cognition*. C&C '24: Creativity and Cognition (Chicago IL USA). ACM, New York, NY, USA, (June 23, 2024). doi:10.1145/3635636.3656201.

[6] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. en. *ACM transactions on computer-human interaction: a publication of the Association for Computing Machinery*, 21, (June 1, 2014), 1–25, 4, (June 1, 2014). doi:10.1145/2617588.

[7] Sandra G Hart. 1986. NASA Task Load Index (TLX): Computerized Version - Volume 1.0, (Jan. 1, 1986). Retrieved Jan. 17, 2025 from https://ntrs.nasa.gov/citations/20000021487.

[8] LazyJot. 2025. Lazyjot. Accessed on 2025-01-20. (2025). https://lazyjot.com/.

[9] Website Learners. 2025. Song lyrics designer. Accessed on 2025-01-20. (2025). https://app.gravitywrite.com/tools/song-lyrics-.

[10] Mina Lee et al. 2024. A Design Space for Intelligent and Interactive Writing Assistants. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (CHI '24) number Article 1054. CHI '24: CHI Conference on Human Factors in Computing Systems (Honolulu HI USA). Vol. 7. ACM, New York, NY, USA, (May 11, 2024), 1–35. ISBN: 9798400703300. doi:10.1145/3613904.3642697.

[11] Rytr LLC. 2025. Song lyrics generator | use cases. Accessed on 2025-01-20. (2025). https://rytr.me/use-cases/song-lyrics/.

[12] Lyrcs. 2024. Lyrcs: Lyrics, Poetry and Rap Writing App. en. Accessed on 2025-01-20, version 3 released on 2024-06-12. (2024). https://lyrcs.app/.

[13] LyricAssistant. 2025. Lyricassistant. Accessed on 2025-01-20. (2025). https://lyricassistant.com/.

[14] Pat Pattison. 1991. *Songwriting: Essential guide to lyric form and structure: Tools and techniques for writing better lyrics*. en. Hal Leonard Corporation, Milwaukee, WI, (Dec. 1, 1991). 118 pp. ISBN: 9781476867533. Retrieved Jan. 20, 2025 from.

[15] Shikaki. 2025. AI Lyrics Writing Tool Shikaki. ja. Accessed on 2025-02-13, latest version released on 2020-11-08. (2025). https://shikaki.diatonic.codes/.

[16] Samangi Wadinambiarachchi, Ryan M Kelly, Saumya Pareek, Qiushi Zhou, and Eduardo Velloso. 2024. The effects of generative AI on design fixation and divergent thinking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. CHI '24: CHI Conference on Human Factors in Computing Systems (Honolulu HI USA). ACM, New York, NY, USA, (May 11, 2024), 1–18. doi:10.1145/3613904.3642919.

[17] Kento Watanabe and Masataka Goto. 2020. Lyrics information processing: analysis, generation, and applications. In *Proceedings of the 1st Workshop on NLP for Music and Audio (NLP4MusA)*. Sergio Oramas, Luis Espinosa-Anke, Elena Epure, Rosie Jones, Mohamed Sordo, Massimo Quadrana, and Kento Watanabe, (Eds.) Association for Computational Linguistics, Online, (Oct. 2020), 6–12. https://aclanthology.org/2020.nlp4musa-1.2/.

[18] Kento Watanabe, Yuichiroh Matsubayashi, Kentaro Inui, Tomoyasu Nakano, Satoru Fukayama, and Masataka Goto. 2017. LyriSys: An Interactive Support System for Writing Lyrics Based on Topic Transition. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (IUI '17) (Limassol, Cyprus). Association for Computing Machinery, New York, NY, USA, (Mar. 7, 2017), 559–563. ISBN: 9781450343480. doi:10.1145/3025171.3025194.

[19] WaveAI Inc. 2024. Lyricstudio - your artificial intelligence ghostwriter. Accessed on 2025-01-20. (2024). https://lyricstudio.net/.

[20] Rongsheng Zhang, Xiaoxi Mao, Le Li, Lin Jiang, Lin Chen, Zhiwei Hu, Yadong Xi, Changjie Fan, and Minlie Huang. 2022. Youling: An AI-assisted lyrics creation system. *arXiv [cs.CL]*, (Jan. 17, 2022). http://arxiv.org/abs/2201.06724 arXiv: 2201.06724 [cs.CL].

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009