PAPER

Kiite Cafe: A Web Service Enabling Users to Listen to the Same Song at the Same Moment While Reacting to the Song*

Kosetsu TSUKUDA†a), **Keisuke ISHIDA**†b), **Masahiro HAMASAKI**†c), ***Nonmembers***,
and* Masataka GOTO**†d), ***Fellow

SUMMARY This paper describes a public web service called *Kiite Cafe* that lets users get together virtually to listen to music. When users listen to music on Kiite Cafe, their experiences are enhanced by two architectures: (i) visualization of each user's reactions, and (ii) selection of songs from users' favorite songs. These architectures enable users to feel social connection with others and the joy of introducing others to their favorite songs as if they were together listening to music in person. In addition, the architectures provide three user experiences: (1) motivation to react to played songs, (2) the opportunity to listen to a diverse range of songs, and (3) the opportunity to contribute as a curator. By analyzing the behavior logs of 2,399 Kiite Cafe users over a year, we quantitatively show that these user experiences can generate various effects (e.g., users react to a more diverse range of songs on Kiite Cafe than when listening alone). We also discuss how our proposed architectures can enrich music listening experiences with others.

key words: *music recommendation, web service, user behavior analysis*

1. Introduction

Unlike listening to music alone, listening to music with others adds the qualities of feeling social connection and letting others listen to one's favorite songs. For example, the former quality occurs when attending a live concert and sharing the experience with other audience members [2], [3], while the latter quality occurs when people introduce others to their favorite songs [4]–[6].

These qualities become hard to enjoy when various social situations or geographic remoteness make it difficult to get together in person and listen to music with others. Instead of attending a live concert, people can listen to the same music at the same time via TV, radio, or live streaming on the web. However, such media provide a poor alternative, because the first quality of social connection requires audiences to get together in the same place so that they can see each other's reactions to the music. Similarly, instead of directly introducing others to favorite songs, people can post URL links (*e.g.*, to YouTube videos of songs) to social networking services (SNSs) such as Twitter and Facebook. However, even if many SNS users react to a song post (*e.g.*, with a "thumbs up"), there is no guarantee that they actually listened to the song and liked it. Rather, the second quality of sharing a favorite song with others requires knowing that people who react actually listened to the song.

In light of the above, we propose a web service called *Kiite Cafe*[**,***], which enables people to get together virtually to listen to music without losing the above qualities. Kiite Cafe is characterized by the following two architectures: (i) when users listen to songs on Kiite Cafe, each user's reactions are visualized; and (ii) songs played on Kiite Cafe are selected from users' favorite songs. To facilitate an intuitive understanding of the user experiences provided by these architectures, we give the following example.

Suppose that Emily is a Kiite Cafe user. One day, she logs in to Kiite Cafe and finds that 39 users are logged in. Each user is identified by his/her own icon. The users, including Emily, can simultaneously listen to the same song, which is automatically selected and played. Even if the played song has a different mood from songs that Emily usually listens to, if she likes it, she can add it to her list of favorite songs (*i.e.*, her *favorites list*). Because she has encountered a new favorite song, she feels happy to listen to a diverse range of songs. Moreover, when the currently played song is added to her favorites list, architecture (i) visualizes her reaction by displaying a heart symbol on her icon. Because other users' reactions are also visualized, she can see their reactions to feel social connection. After a short time, one of Emily's favorite songs is played when it is automatically selected by architecture (ii). While her favorite song is playing, she is pleased to notice that a heart symbol is displayed on another user's icon. Then, other users also react to the song, and eventually the heart symbol is displayed on eight users' icons. Architecture (i) thus enables Emily to see the moments when other users start liking one of her favorite songs. This experience, in which she contributes as a curator[****], makes her feel happy and want other users to listen to another of her favorite songs. Thus, Emily looks forward to another favorite song being played; until then, she stays on Kiite Cafe and enjoys other users' favorite songs.

Manuscript received January 6, 2023.
Manuscript revised May 26, 2023.
Manuscript publicized July 28, 2023.

[†] The authors are with National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba-shi, 305–8568 Japan.

[*] The original version of this paper was published at a conference [1]. We have extended the paper by conducting experiments with logs covering a longer period and by discussing more application examples for online events.

a) E-mail: k.tsukuda@aist.go.jp
b) E-mail: ksuke-ishida@aist.go.jp
c) E-mail: masahiro.hamasaki@aist.go.jp
d) E-mail: m.goto@aist.go.jp

DOI: 10.1587/transinf.2023EDP7001

[**] "Kiite" means "Listen" in Japanese.

[***] https://cafe.kiite.jp

[****] In this paper, we use the word "curator" for a person who introduces or shares a song in a way that adds value.

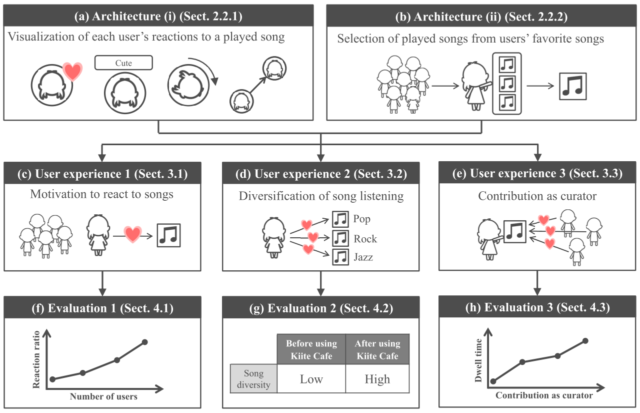

(a) Architecture (i) (Sect. 2.2.1)
Visualization of each user's reactions to a played song
Cute
(b) Architecture (ii) (Sect. 2.2.2)
Selection of played songs from users' favorite songs
(c) User experience 1 (Sect. 3.1)
Motivation to react to songs
(d) User experience 2 (Sect. 3.2)
Diversification of song listening
Pop
Rock
Jazz
(e) User experience 3 (Sect. 3.3)
Contribution as curator
(f) Evaluation 1 (Sect. 4.1)
Reaction ratio
Number of users
(g) Evaluation 2 (Sect. 4.2)
Before using Kiite Cafe
After using Kiite Cafe
Song diversity
Low
High
(h) Evaluation 3 (Sect. 4.3)
Dwell time
Contribution as curator

Fig. 1 Structure of this paper.

Our contributions in this work, as illustrated in Fig. 1, can be summarized as follows.

- We propose two architectures for enabling people to simultaneously listen to the same music online while achieving the qualities of social connection and the joy of introducing other people to favorite songs (Fig. 1-(a) and (b)).
- We implemented and released a web service, called Kiite Cafe, that applies these architectures.
- We describe three user experiences in which users (1) are motivated to react to songs, (2) can listen to a diverse range of songs, and (3) can contribute as curators; we also discuss the effects of these experiences on users as a result of the two proposed architectures (Fig. 1-(c), (d), and (e)).
- By analyzing logs of user behavior on Kiite Cafe, we show that the architectures do provide the above effects. Specifically, users (1) react to songs more actively as the number of users on Kiite Cafe increases, (2) react to a more diverse range of songs on Kiite Cafe than when they listen to songs alone, and (3) stay on Kiite Cafe longer when they contribute more as curators (Fig. 1-(f), (g), and (h)).

2. Overview of Kiite Cafe

Kiite Cafe is implemented as a novel function on Kiite[†], which is an existing web service for exploring and discovering music. Any Kiite user can use Kiite Cafe. Below, we introduce the Kiite functions that are related to Kiite Cafe and then give an overview of Kiite Cafe.

2.1 Kiite

Song data on Kiite are routinely collected from *Nico Nico Douga*[††], which is one of the most popular video sharing services in Japan. On Nico Nico Douga, it is quite popular for both amateur and professional musicians to upload songs created with singing voice synthesizer software called *VOCALOID* [7]. As of the end of December 2022, more than 410,000 songs can be played back on Kiite. When a Kiite user listens to a song, its video clip is played on Kiite by an embedded video player[†††].

Kiite enables users to effectively find favorite songs by providing novel functions such as exploration of songs based on their impressions and continuous listening to only the choruses of multiple songs. A registered user can set her own icon image, add songs to her favorites list, create playlists, listen to other users' playlists, and so on.

2.2 Kiite Cafe

Figure 2 shows an overview of Kiite Cafe. When a user logs in, her icon is displayed at a random position in a two-dimensional space that also displays other logged-in users. All of the users listen to the same song played in a video player (Ⓐ in the figure) at the same time. As mentioned in Sect. 1, Kiite Cafe has two architectures, for visualizing users' reactions and selecting songs to play from users' favorite songs. In the rest of this section, we describe the details of each architecture.

2.2.1 Architecture (i): User Reaction Visualization (Fig. 1-(a))

We visualize the following four kinds of reactions so that users can see each other's reactions to a played song.

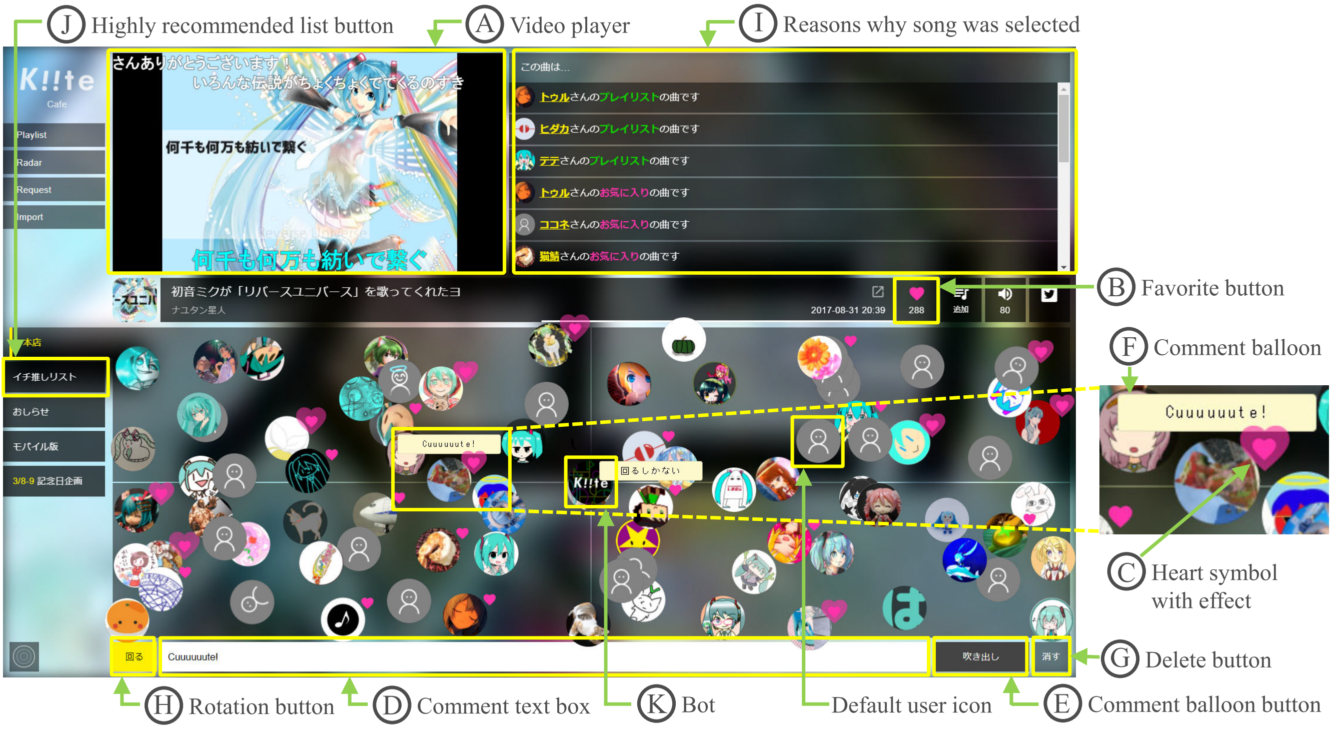
Favorite. When a user likes a played song, she can add it to her favorites list by clicking the "favorite" button Ⓑ. When the button is clicked, a heart symbol with an animation effect is displayed at the top right of the user's icon while the song is playing (Ⓒ). This enables users to quickly see how many users like a song. When the user had already added the played song to her favorites list, the heart symbol is displayed without the effect.

Comment. When a comment is entered in a text box Ⓓ and the "comment balloon" button Ⓔ is clicked, a comment balloon is displayed above the user's icon for 90 seconds (Ⓕ). The user can also manually delete her comment by clicking the "delete" button Ⓖ. Users can thus use this function to express their impressions of a played song or have simple communication with each other.

Rotation. A user can rotate her icon by clicking the "rotation" button Ⓗ. The icon then rotates clockwise at a uniform rate until the played song ends. The user can also manually stop the rotation by clicking the "rotation" button again. Users can use this function to express feelings like a sense of excitement. However, note that Kiite Cafe does not provide any guidance on when users should use this function, because we want them to use it as they please.

Move. By clicking an arbitrary position in the two-dimensional space, a user can move her icon to the clicked position. The icon is animated to move to the position in

[†] https://kiite.jp
[††] https://www.nicovideo.jp
[†††] On Nico Nico Douga, all songs are uploaded as music videos.



この曲は…

トウルさんのプレイリストの曲です

ヒダカさんのプレイリストの曲です

テテさんのプレイリストの曲です

トウルさんのお気に入りの曲です

ココネさんのお気に入りの曲です

猫鯖さんのお気に入りの曲です

Fig. 2 Screenshot of Kiite Cafe.

a straight line at a uniform rate. Kiite Cafe does not display any meaning for the quadrants and axes in the two-dimensional space. Instead, as with the *Rotation* function, we leave the usage of the *Move* function to users.

2.2.2 Architecture (ii): Song Selection from Users' Favorite Songs (Fig. 1-(b))

Let U denote a set of users who are logged in to Kiite Cafe. For each user u, we define S_u as the set of songs included in u's favorites list or playlists. A played song is selected from $\bigcup_{u \in U} S_u$. The automatic song selection process is invoked before the end of the currently played song, and it consists of the following two steps: (1) selection of a user, and (2) selection of a song from the user's favorite songs.

In the first step, if there are biases toward certain selected users, then the selected songs may also be biased. Moreover, some users may become frustrated if their favorite songs are not selected at all. To avoid such biases and satisfy every user, we developed an algorithm that can randomly but fairly select users and thus diversify the played songs. Suppose that user u is selected in the first step, such that every user has an equal chance to be selected. When song $s \in S_u$ is randomly selected in the second step, the reason for its selection is displayed, *e.g.*, "This song is in u's playlist" (first row of Ⓘ in Fig. 2). If s is also among other users' favorite songs, that information is displayed (second and later rows of Ⓘ) so that those users can notice that one of their favorite songs is being played. Moreover, a user can set one of her playlists as a "highly recommended list" by clicking a button Ⓙ. When the selected user sets a list as "highly recommended," a song in that list is randomly but preferentially selected in the second step. By setting such a list, a user can specify the songs that she wants other users to listen to.

Note that the implemented selection algorithm described above for our service is just an example, and other algorithms can be used as long as they balance the fairness and randomness of selecting both users and songs. That is, the algorithm itself is not an essential part of this paper. Rather, the essential parts are the two proposed architectures and the new music listening experiences that they provide (see Sect. 3).

In addition, we created a bot account Ⓚ that is always logged in. The bot periodically creates playlists according to a daily/weekly popularity ranking of VOCALOID songs on Nico Nico Douga. The bot is treated as one of the users, and songs in its playlists can also be selected by the song selection process. This gives a user a chance to listen to the latest popular songs and find new favorite songs even when no other human users are logged in. Note that the bot does not show any reactions to played songs.

3. User Experiences and Effects

As mentioned in Sect. 1, the proposed Kiite Cafe architectures add two qualities: social connection, and the joy of introducing others to favorite songs. In addition, the architectures provide three kinds of user experiences. This section describes those experiences and their effects on users.

3.1 Motivation to React to Songs (Fig. 1-(c))

Although many studies have been conducted on enabling users to listen to music together, most of them have focused on visualizing the song selection process or proposing methods for that process [8]–[12]. A system that can show a summary of listeners' feedback on a song (total numbers of likes and dislikes) has been proposed [13]; however, little attention has been paid to visualizing each user's reactions. In contrast, Kiite Cafe visualizes users' reactions via their icons, as described in Sect. 2.2.1. By sharing all the users' reactions with each other, Kiite Cafe motivates them to react to the currently played song. Accordingly, we expect that, the more people get together on Kiite Cafe, the more meaningful it will be to show their reactions, and the more actively they will react to songs. In the long term, this would enable users to develop the habit of actively listening to music and enrich their listening experiences [14].

3.2 Diversification of Song Listening (Fig. 1-(d))

Many studies have sought to play songs that match the musical preferences of as many users as possible [8], [10], [15]. In the short term, this approach may be able to increase users' satisfaction. In the long term, however, as is known from the negative effects of a filter bubble [16], [17], this approach could narrow users' musical interests. On the other hand, because Kiite Cafe plays a diverse range of songs selected from various users' favorite songs, it enables users to find not only songs that match their preferences well but also unexpected or serendipitous songs [18] that do not match their usual preferences. In other words, we expect that a user will react to a more diverse range of songs on Kiite Cafe than when she listens alone. In the long term, this experience would expand the user's horizons.

Of course, it may happen that some songs played on Kiite Cafe do not match a user's musical preferences, and the user may not like them even after listening to them. However, by diversifying the selected songs, it is unlikely that only songs the user does not like are played all the time. This motivates users to look forward to what song will be played next, even if a song that does not match their preferences is played.

3.3 Contribution as Curators (Fig. 1-(e))

According to architecture (ii), suppose that a song in user u's playlist is selected and played on Kiite Cafe. Because of architecture (i), u can see the moment when other users start liking or show interest in that song (*e.g.*, u can see when other users add the song to their favorites list or rotate their icon). For other users, u effectively plays a role as a curator. That is, the two architectures enable every user to naturally contribute as a curator. We expect that when a user experiences the joy of contributing as a curator, she will look forward to the curation opportunity when another of her favorite songs is played and thus increase her dwell time on Kiite Cafe. Acting as a curator has been reported to increase music listening activity (*e.g.*, listening to more songs and making playlists for curation) [19]. Therefore, in the long term, this experience would promote users' daily music listening activity.

4. Experiment

We officially launched the Kiite Cafe service on August 5, 2020. In this section, we evaluate the three expected effects discussed in the previous section. To this end, we analyzed user behavior logs for the period between August 5, 2020 and August 4, 2021. The number of unique users who logged in during this period was 2,399. The *Favorite*, *Comment*, *Rotation*, and *Move* reactions were used 49,425, 14,094, 109,054, and 74,304 times, respectively.

4.1 Frequency of User Reactions (Fig. 1-(f))

As a result of users sharing their reactions with each other on Kiite Cafe, we expect that they will be more motivated to react as the number of users increases (Sect. 3.1). To verify this effect, we evaluated the following research question: *Does a user react to a played song more frequently as the number of users on Kiite Cafe increases? (**RQ1**)*

4.1.1 Settings

We considered the four kinds of reactions: $R = \{Favorite, Comment, Rotation, Move\}$. First, for each played song, we obtained U_s, the set of users except the bot who were on Kiite Cafe when song s started playing. According to the number of users (*i.e.*, $|U_s|$), we categorized songs into four classes (C_1: $1 \leq |U_s| \leq 5$; C_2: $6 \leq |U_s| \leq 10$; C_3: $11 \leq |U_s| \leq 15$; C_4: $16 \leq |U_s|$). To answer **RQ1**, for each reaction, we compared the average proportion of users who reacted to a song among the classes.

More formally, let S_{C_i} denote a list of songs in C_i $(1 \leq i \leq 4)$[†]. Given song $s \in S_{C_i}$ and reaction $r \in R$, let U_s^r denote the set of users who gave r as a reaction to s. Then, the proportion of such users is given by $ratio(s, r) = \frac{|U_s^r|}{|U_s|}$. Finally, the average proportion over S_{C_i} was computed as follows.

$$avgratio(S_{C_i}, r) = \frac{1}{|S_{C_i}|} \sum_{s \in S_{C_i}} ratio(s, r). \qquad (1)$$

4.1.2 Results

Figure 3 shows the results. For visibility, $avgratio(S_{C_i}, r)$ was normalized by $avgratio(S_{C_1}, r)$ for each reaction. All of the reaction proportions monotonically increased as the number of users increased; thus, the answer to **RQ1** is

[†] Because the same song can be played multiple times on Kiite Cafe, the same song can appear multiple times in S_{C_i}.

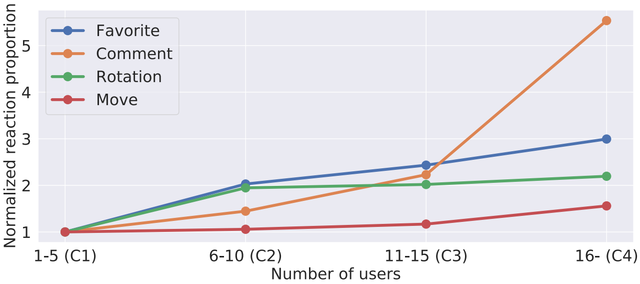
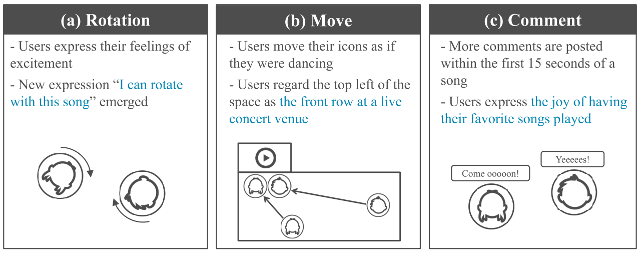


Fig. 3 Relation between the number of users on Kiite Cafe and the normalized reaction proportion.



(a) Rotation

- Users express their feelings of excitement

- New expression "I can rotate with this song" emerged

(b) Move

- Users move their icons as if they were dancing

- Users regard the top left of the space as the front row at a live concert venue

(c) Comment

- More comments are posted within the first 15 seconds of a song

- Users express the joy of having their favorite songs played

Fig. 4 Characteristic usages of "Rotation," "Move," and "Comment."

"Yes." Because the *Favorite* function was obviously used to add a song to a user's favorites list, we discuss how the users used the other three functions, as illustrated in Fig. 4.

Regarding the *Rotation* function, although Kiite Cafe does not explain its purpose, we searched Kiite Cafe users' tweets on Twitter[†] and found that a number of users used this function to express their feelings of excitement. As a result of this usage spreading, a new expression "I can rotate with this song" emerged on Kiite Cafe. We expect this expression to spread in music culture in the future (Fig. 4-(a)).

Next, by analyzing tweets about the *Move* function, we found that it was used mainly for two purposes. First, users moved their icons as if they were dancing. Second, users regarded the top left (or the top) of the two-dimensional space (*i.e.*, near the video player) as the front row at a live concert venue, and they moved there when their favorite songs were played (Fig. 4-(b)). It is interesting that such a culture was created by the users and spread among them.

Regarding the *Comment* function, although the average length of the played songs was 234 seconds, 9.91% of comments were posted within the first 15 seconds of a song. In such comments, users often expressed the joy of having their favorite songs played (*e.g.*, "Come oooon!" and "Yeeeees!"). This was similar to the phenomenon at live concerts in which the audience gets excited when a favorite song starts (Fig. 4-(c)). In summary, as the number of users increased, they were more likely to express their excitement and behave as if they were attending a live concert.

Unlike the *Favorite* and *Comment* functions, the *Rotation* and *Move* functions allow users to express their impressions for songs more casually. Since the *Favorite* function adds a song to the favorites list, users may not use its function unless they really like the song being played. Although the *Comment* function is versatile, users may find it burdensome to enter text. In fact, as described at the beginning of Sect. 4, while the *Favorite* and *Comment* functions were used 49,425 and 14,094 times, respectively, the *Rotation* and *Move* functions were used as many as 109,054 and 74,304 times, respectively.

4.2 Diversity of Reacted Songs (Fig. 1-(g))

Because Kiite Cafe enables users to listen to songs that do not always match their musical preferences, we expect that they will react to a more diverse range of songs (Sect. 3.2). To verify this effect, we evaluated the following research question: *Does a user react to a more diverse range of songs on Kiite Cafe as compared to her musical preferences before she started using the service? (**RQ2**)*

4.2.1 Settings

Let t_u denote the time when user u initially logged in to Kiite Cafe. We assumed that songs added to u's favorites list before t_u (*i.e.*, before using Kiite Cafe), denoted by S_u^{org}, represented u's original musical preferences. These songs were collected on the original Kiite service, which was launched on August 30, 2019, as described in Sect. 2.1. Moreover, we assumed that songs for which u gave reaction r, denoted by S_u^r, represented u's musical preferences in terms of r after starting to use Kiite Cafe. To answer **RQ2**, we compared the diversity of S_u^r with that of S_u^{org} for each reaction.

Formally, the diversity was computed as the intra-list diversity [20]. In the case of S_u^{org},

$$div(S_u^{org}) = \frac{\sum_{s_i \in S_u^{org}} \sum_{s_j \in S_u^{org} \setminus \{s_i\}} dist(s_i, s_j)}{|S_u^{org}|(|S_u^{org}| - 1)}, \tag{2}$$

where $dist(s_i, s_j)$ is the Euclidean distance between the audio-based feature vectors [21] of s_i and s_j. Those vectors were one of the state-of-the-art features, though arbitrary suitable audio-based features could also be used. For each reaction r, to appropriately measure users' musical preferences, we considered only users who had more than nine songs in both S_u^{org} and S_u^r[††]. Let U^r denote the set of such users. Then, given r, the average diversities of S_u^{org} and S_u^r were computed as follows:

$$avgdiv(S_u^{org}) = \frac{1}{|U^r|} \sum_{u \in U^r} div(S_u^{org}), \tag{3}$$

$$avgdiv(S_u^r) = \frac{1}{|U^r|} \sum_{u \in U^r} div(S_u^r). \tag{4}$$

[†] We assumed that Twitter users who tweeted about the function were Kiite Cafe users.

[††] Because we released a beta version of Kiite Cafe on May 1, 2020, users who logged in to Kiite Cafe for the first time between May 1, 2020 and August 4, 2020 were not included in this analysis.

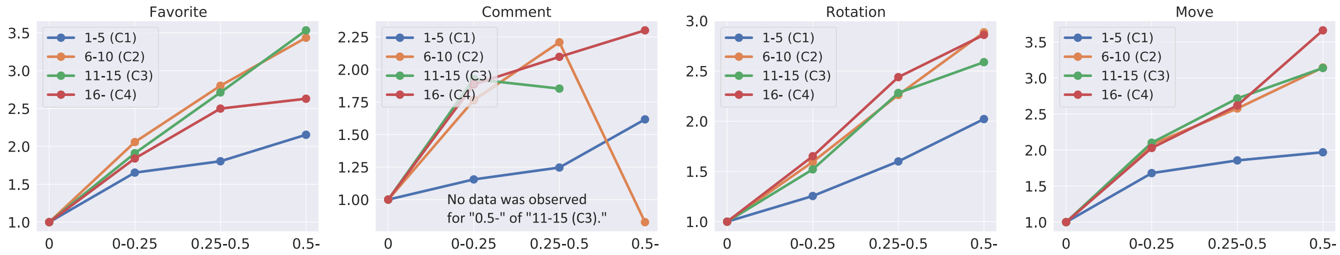


Fig. 5 Relations between the proportion of users who gave reactions (x-axis) and the normalized dwell time (y-axis).

Table 1 Diversity of musical preferences.

Reaction r	$\lvert U^r\rvert$	$avgdiv(S_u^{org})$	$avgdiv(S_u^r)$	p-value
Favorite	157	10.509	10.994	1.14×10^{-8}
Comment	69	10.454	10.954	2.05×10^{-3}
Rotation	143	10.519	10.895	1.16×10^{-6}
Move	131	10.454	11.059	2.75×10^{-9}

4.2.2 Results

Table 1 lists the results. A paired t-test was conducted under the assumption that the distribution of diversity scores before and after starting to use Kiite Cafe follows a normal distribution for each. We found that, for all reactions, the diversity of songs producing reactions statistically significantly increased in comparison to the diversity of favorite songs before a user started to use Kiite Cafe; thus, the answer to **RQ2** is "Yes." These results indicate that Kiite Cafe is also useful as a service for users to find songs that are different from their daily musical preferences.

Furthermore, to be sure that this increase in diversity was not due to an increase in the diversity of VOCALOID songs uploaded during the same period, we similarly compared the diversities using only VOCALOID songs prior to the release of Kiite Cafe on August 5, 2020. The results showed the same trend as in Table 1, with a statistically significant increase in the diversity for all the four kinds of reactions.

4.3 Dwell Time (Fig. 1-(h))

Because Kiite Cafe enables users to experience the joy of contributing as curators, we expect that they will stay longer as their contributions increase (Sect. 3.3). To verify this, we evaluated the following research question: *Does a user stay on Kiite Cafe for a longer time as the proportion of users who react to her favorite songs increases? (**RQ3**)*

4.3.1 Settings

We define user u's session on Kiite Cafe as the duration between u's login and logout times. In u's kth session, suppose that three of u's favorite songs were played, and that 0%, 40%, and 16% of users gave reaction r to those songs. Following the assumption that the maximum percentage (in this case, 40%) influenced u's dwell time, we categorized the maximum value of $ratio(s, r)$, as defined in Sect. 4.1.1, into four classes (G_1: $ratio(s, r) = 0$; G_2: $0 < ratio(s, r) \leq 0.25$; G_3: $0.25 < ratio(s, r) \leq 0.5$; G_4: $0.5 < ratio(s, r)$). However, for a proportion of 0.4, eight reacting users among 20 users would have a higher impact on u than two reacting users among five users. Therefore, we also considered the classes of $|U_s|$ (*i.e.*, C_i) as defined in Sect. 4.1.1. That is, to answer **RQ3**, given a class of the number of users, we compared the average session lengths between the reaction proportion classes for each reaction.

Formally, let D_u and S_{G_j} denote a list of u's sessions and a list of songs in G_j ($1 \leq j \leq 4$), respectively. $T_{u,k} \in D_u$ represents a list of songs from u's favorite songs (S_u) that were played in the kth session. For that session, we selected the song $s^{max}_{u,k,r} \in T_{u,k}$ that had the highest proportion of users who gave reaction r (*i.e.*, $s^{max}_{u,k,r} = \arg\max_{s \in T_{u,k}} ratio(s, r)$). Given C_i and G_j, we define the set of $s^{max}_{u,k,r}$ belonging to S_{C_i} and S_{G_j} in all users' sessions as $S_{i,j,r} = \{s^{max}_{u,k,r} \mid u \in U \land 1 \leq k \leq |D_u| \land s^{max}_{u,k,r} \in T_{u,k} \land s^{max}_{u,k,r} \in S_{C_i} \land s^{max}_{u,k,r} \in S_{G_j}\}$. Let $len(u, k)$ denote the length in seconds of u's kth session. Then, the average session length was computed as

$$avglen(S_{i,j,r}) = \frac{1}{|S_{i,j,r}|} \sum_{s^{max}_{u,k,r} \in S_{i,j,r}} len(u, k). \tag{5}$$

4.3.2 Results

Figure 5 shows the results; for visibility, $avglen(S_{i,j,r})$ was normalized by $avglen(S_{i,1,r})$ for each reaction. For the *Favorite*, *Rotation*, and *Move* functions, we can see that the dwell time tended to increase as the proportion of users who gave that reaction increased. In these graphs, the line for the class of 1-5 users (C_1) is located at the lowest position among the four classes (C_1 - C_4). These results indicate that not only the proportion of users who gave a reaction but also the absolute number of such users influenced the dwell time. On the other hand, because the frequency of comments was lower than the frequencies of the other reactions, no clear tendency was observed for the *Comment* function when the number of users was 6-10 (C_2) or 11-15 (C_3). The significantly reduced dwell time for "0.5-" section of 6-10 (C_2) appears to be an outlier, but this is because there were only three corresponding session data. Because the dwell time monotonically increased when the number of users was 1-5

(C_1) or at least 16 (C_4), the *Comment* function could also potentially have a positive effect. Detailed analysis with more user behavior logs will be required to verify this effect, and we leave that for a future work. In summary, the answer to **RQ3** is “Yes” for *Favorite*, *Rotation*, and *Move*.

5. Discussion

In Sect. 3, we described the user experiences provided by Kiite Cafe and the effects of those experiences. We believe that Kiite Cafe has even more potential to diversify and enrich users’ music listening experiences. In this section, to demonstrate that potential, we discuss four themes.

5.1 Application Examples for Online Events

Kiite Cafe has been used for several online events. During an event, instead of using architecture (ii) to select the songs to play, the participants listened to a designated playlist together while communicating with each other via architecture (i). The events are classified into the following three categories, depending on the originator.

The first category is events initiated by the organizers of VOCALOID-related events. Thus far, these events have been held in the form of collaborations with VOCALOID-related events such as “Creators Market Online”[†], “Tsunagaru Mirai”[††], “Tamesareru Mirai”[†††], and “VOCALOID DJ STATION”[††††]. At an event related to “Creators Market Online” on August 29, 2020, for example, a famous creator of VOCALOID songs made a special playlist that consisted of songs that the creator liked or had created. During the one-hour event, as many as 140 Kiite Cafe users enjoyed simultaneously listening to the songs in the playlist, and they used the reaction functions of Kiite Cafe to communicate with the creator in real time. For another event related to “Tamesareru Mirai” on February 11, 2021, on their web page, a questionnaire was conducted on favorite VOCALOID songs related to winter or snow. During the 90-minute event, 77 users enjoyed listening to songs in a playlist created according to the questionnaire answers.

The second category is events initiated by the users of Kiite Cafe. Since March 19, 2021, a link “Anniversary event wanted” has been displayed on Kiite Cafe. The link brings up a form for a user to enter information about the user’s desired event, such as the name of the singing voice synthesizer character, the type of anniversary event, and the event date. The event to be held is determined from these user proposals by considering the number of proposals, schedule, feasibility, and so on. A playlist to be played during a character’s event is created according to certain statistics such as the number of favorites of each song by the character on Kiite Cafe. Events such as the “Kasane Teto anniversary event” on April 1, 2021 and the “GUMI anniversary event” on June 26, 2021 were held, and many users participated. These dates were the characters’ birthdays, and each event took place around midnight (23:30 to 0:15). At the stroke of midnight, many users posted celebratory comments such as “Happy Birthday!” while enjoying the music.

The third category is events initiated by the management team of Kiite Cafe. For these events, the playlists were created in the same way as in the second category. For example, on March 9, 2021, the “Hatsune Miku anniversary event” was held as a token of “Miku”[†††††]. On August 31, 2021, the “Hatsune Miku birthday event” was held and was enjoyed by as many as 285 users.

Although it has become difficult for people to get together in person and communicate with each other and with creators because of the COVID-19 pandemic, we have demonstrated a new style of online music events through these examples. Moreover, even after the pandemic’s resolution, we believe that this kind of online event will be valuable for users who cannot easily attend physical events for reasons such as geographic remoteness.

5.2 Additional Service Functions

Although all users on Kiite Cafe get together in one online space, it would be interesting to provide additional spaces for different purposes in the future: we could call the main space and additional spaces the “main cafe” and “branches,” respectively. For example, for a branch on the theme of “time,” we could put a higher priority on songs related to time (*e.g.*, playing night-related songs at night) by analyzing song lyrics if they are available in the song selection process.

We could also consider a function that allows any user to conduct a questionnaire by displaying possible responses in each quadrant of the two-dimensional space. For example, a user might ask “Who would you like to listen to the played song with?” and assign responses of “family,” “lover,” “friend,” and “other” to the quadrants. Other users could answer this question by moving their icons. This function would provide a good opportunity to see how other users perceive a song.

5.3 Trustworthiness of Kiite Cafe

With the recent spread of artificial intelligence (AI) technology, the concept of “trustworthy AI” has become increasingly important to foster “trust” that is essential for the acceptance of AI technology in society [22]–[24]. The concepts of Kiite Cafe can also contribute to the realization of trustworthy AI as follows. First, because songs played on Kiite Cafe depend on users’ favorite songs, they are diverse

[†] https://karent.jp/creatorsmarket

[††] http://tsunagarumirai.com/2020

[†††] https://twipla.jp/events/472240

[††††] https://vdds-official.tumblr.com

[†††††] In Japanese, 3 and 9 can be pronounced as “mi” and “ku,” respectively. On March 9, 2021, 140 Kiite Cafe users were enjoying the event at the same time, and two years later, on March 9, 2023, 754 users were enjoying the event at the same time, indicating the growing popularity of this service.

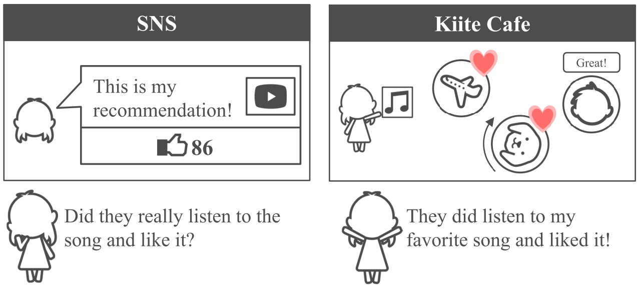

SNS
This is my recommendation!
86
Did they really listen to the song and like it?
Kiite Cafe
Great!
They did listen to my favorite song and liked it!

Fig. 6 Comparison between an SNS and Kiite Cafe.

and are less likely to be biased (*e.g.*, bias such as songs created by a particular creator are frequently played). Moreover, as we described in Sect. 2.2.2, our algorithm fairly selects users. Therefore, *fairness* in terms of both played songs and selected users is realized on Kiite Cafe. Second, a user could keep track of the recommended songs on Kiite Cafe and verify that the fairness of the songs being selected is truly satisfied. Therefore, Kiite Cafe realizes *transparency* in that an external observer can verify the recommendation operation. Third, *explainability* is guaranteed because the reasons why a played song was selected are displayed to users as we explained in Sect. 2.2.2 and Fig. 2 ①.

5.4 Reusable Insights

The reusable insights of the work can be summarized as follows.

- Through the experiments, we verified that the two architectures are effective in promoting users' music listening activity. These architectures can be helpful for other researchers and companies to develop interfaces that enable users to listen to music together.
- The examples of successful online events showed that the architectures can offer new ways to enjoy music with other people even in situations where it is difficult for people to physically get together in the same place. We thus opened up a new research theme to support interactions among creators, audiences, and music.
- We clarified the value of visualizing the moment when a user starts liking a song. In contrast to traditional curation on an SNS, the Kiite Cafe approach guarantees that users who liked a user's favorite song actually listened to it, as shown in Fig. 6. This insight could also be beneficial in designing other music listening systems or services.

6. Related Work

6.1 Music Listening Systems for Group of Users

Music listening systems for a single user were reviewed by Goto and Dannenberg [25] and by Knees et al. [26]. In contrast, systems for a group of users can be classified into two types. The first type aims to enable users to listen to music at the same time. Most studies on this type assume that users get together in person at a public space such as a fitness center [8], a party [11], a bar [10], or a room [9]. In MusicFX [8] and Flytrap [9], the system reads users' musical preferences from each user's device, and songs stored in the system are played by taking those preferences into account. In contrast, in Jukola [10], PartyVote [11], and WePlay [13], users nominate songs to be played, like a jukebox. In the second type of group listening system, users share songs with other users. Sharing music with others is an important activity to expand listeners' horizons [5]. Studies on this type do not assume that users listen to a song at the same time. Push!Music [5] and tunA [4] are mobile music players that let users share songs via Wi-Fi with others who are nearby. The user studies on those systems showed that users are comfortable sharing their favorite songs with others whether they are friends or strangers. It has also been reported that users share songs mainly because they want to recommend songs that others would like, disseminate their favorite songs, talk about shared songs with others, and so on.

Some applications designed for listening to music together have also been released (*e.g.*, Group Session by Spotify [27] and JQBX [28]). In these applications, any user can let other users listen to her favorite songs by acting like a DJ. Users can also communicate with each other via a text chat system while listening to songs.

Our study is different from the above studies and services in that we introduce the two architectures for reaction visualization and song selection from users' favorite songs. In most of the above cases, because users' reactions are not visualized or are visualized only when chatting with text messages, it is difficult for users to feel social connection with each other. On the other hand, because the first architecture on Kiite Cafe visualizes four kinds of reactions, users can more strongly feel that they are enjoying music with others. In addition, existing systems require users to actively nominate or share songs or act like a DJ, but some users may hesitate to do that, especially if there is a large audience. In contrast, the second architecture on Kiite Cafe enables a user's favorite songs to be automatically played. This lets any user share her favorite songs with other users and see the moments when they start liking those songs.

6.2 Group Recommendation Algorithms

Various song recommendation methods for a single user have been proposed [29]–[36]. One of the biggest differences between the methods for a single user and those for a group of users is that the latter methods need to take multiple users' preferences into account. A general approach is to aggregate each user's preferences by, for example, merging recommendation results generated for each user according to voting strategies [15], [37]. However, such an approach cannot always reflect minority preferences.

To solve this problem, a concept of *fairness* has recently been introduced into group recommendation algorithms [38]–[42]. The basic idea of fairness is that a list of items recommended to a group is fair when each user in the group can find at least one item in the list that she finds satisfying. In the context of music recommendation, existing studies have only considered the fairness for users as audiences. On the other hand, Kiite Cafe achieves fairness for users as both curators and audience members because of the second architecture, in which each user's favorite songs are fairly selected and played as described in Sect. 2.2.2. In particular, the "highly recommended list" plays an important role in achieving fairness for users as curators. When a user's favorite and/or recommended song can be listened to with other users, the user is satisfied from both the audience and curator viewpoints.

7. Conclusion

In this paper, we described Kiite Cafe, a web service that enables users to communicate while listening to the same music online. Kiite Cafe is characterized by two proposed architectures for visualizing each user's reactions and selecting played songs from users' favorite songs. Our experimental results quantitatively showed three effects of the proposed architectures. Since VOCALOID songs are already diverse and the proposed architectures are not limited to VOCALOID songs, we expect them to be effective for a wide variety of music. We believe that these architectures are also useful for different types of music interfaces, including a three-dimensional interface in which user avatars can listen to the same music in a virtual reality (VR) venue.

Acknowledgments

We thank Crypton Future Media, INC. for developing Kiite with us and Nico Nico Douga for initially tacitly (later explicitly) encouraging us to work on this research. We also would like to extend our appreciation to users of Kiite and Nico Nico Douga, creators of VOCALOID songs, creators of popularity rankings of VOCALOID songs, and all people who have created, supported, and enjoyed the VOCALOID culture.

This work was supported in part by JST CREST Grant Number JPMJCR20D4, JST ACCEL Grant Number JPMJAC1602, and JSPS KAKENHI Grant Numbers JP21H04917 and JP20K19934.

References

[1] K. Tsukuda, K. Ishida, M. Hamasaki, and M. Goto, "Kiite Cafe: A web service for getting together virtually to listen to music," Proc. 22nd International Society for Music Information Retrieval Conference, ISMIR 2021, pp.697–704, 2021.

[2] K. Sedgman, "Coughing and clapping: Investigating audience experience," Cultural Trends, vol.24, no.4, pp.324–326, 2015.

[3] S.C. Brown and D. Knox, "Why go to pop concerts? The motivations behind live music attendance," Musicae Scientiae, vol.21, no.3, pp.233–249, 2017.

[4] A. Bassoli, J. Moore, S. Agamanolis, and H.C. Group, "tunA: Local music sharing with handheld Wi-Fi devices," Proc. 5th Wireless World Conference, WWC 2004, pp.1–23, 2004.

[5] M. Håkansson, M. Rost, and L.E. Holmquist, "Gifts from friends and strangers: A study of mobile music sharing," Proc. 10th European Conference on Computer-Supported Cooperative Work, ECSCW 2007, pp.311–330, 2007.

[6] M. Håkansson, M. Rost, M. Jacobsson, and L.E. Holmquist, "Facilitating mobile music sharing and social interaction with Push!Music," Proc. 40th Annual Hawaii International Conference on System Sciences, HICSS 2007, pp.87–96, 2007.

[7] H. Kenmochi and H. Ohshita, "VOCALOID - commercial singing synthesizer based on sample concatenation," Proc. 8th Annual Conference of the International Speech Communication Association, INTERSPEECH 2007, pp.4009–4010, 2007.

[8] J.F. McCarthy and T.D. Anagnost, "MusicFX: An arbiter of group preferences for computer supported collaborative workouts," Proc. 1998 ACM Conference on Computer Supported Cooperative Work, CSCW 1998, pp.363–372, 1998.

[9] A. Crossen, J. Budzik, and K.J. Hammond, "Flytrap: Intelligent group music recommendation," Proc. 7th International Conference on Intelligent User Interfaces, IUI 2002, pp.184–185, 2002.

[10] K. O'Hara, M. Lipson, M. Jansen, A. Unger, H. Jeffries, and P. Macer, "Jukola: Democratic music choice in a public space," Proc. 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS 2004, pp.145–154, 2004.

[11] D. Sprague, F. Wu, and M. Tory, "Music selection using the PartyVote democratic jukebox," Proc. Working Conference on Advanced Visual Interfaces, AVI 2008, pp.433–436, 2008.

[12] G. Popescu and P. Pu, "What's the best music you have?: Designing music recommendation for group enjoyment in GroupFun," Proc. CHI '12 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2012, pp.1673–1678, 2012.

[13] F. Vieira and N. Andrade, "Evaluating conflict management mechanisms for online social jukeboxes," Proc. 16th International Society for Music Information Retrieval Conference, ISMIR 2015, pp.190–196, 2015.

[14] M. Goto, "Active music listening interfaces based on signal processing," Proc. 2007 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2007, pp.IV-1441–IV-1444, 2007.

[15] M. Kompan and M. Bielikova, "Group recommendations: Survey and perspectives," Computing & Informatics, vol.33, no.2, 2014.

[16] E. Pariser, The filter bubble: What the Internet is hiding from you, Penguin Press, 2011.

[17] M. Taramigkou, E. Bothos, K. Christidis, D. Apostolou, and G. Mentzas, "Escape the bubble: Guided exploration of music preferences for serendipity and novelty," Proc. 7th ACM Conference on Recommender Systems, RecSys 2013, pp.335–338, 2013.

[18] Y.C. Zhang, D.Ó. Séaghdha, D. Quercia, and T. Jambor, "Auralist: Introducing serendipity into music recommendation," Proc. 5th ACM International Conference on Web Search and Data Mining, WSDM 2012, pp.13–22, 2012.

[19] J. Fuller, L. Hubener, Y. Kim, and J.H. Lee, "Elucidating user behavior in music services through persona and gender," Proc. 17th International Society for Music Information Retrieval Conference, ISMIR 2016, pp.626–632, 2016.

[20] C.-N. Ziegler, S.M. McNee, J.A. Konstan, and G. Lausen, "Improving recommendation lists through topic diversification," Proc. 14th International Conference on World Wide Web, WWW 2005, pp.22–32, 2005.

[21] A.L. Cramer, H.-H. Wu, J. Salamon, and J.P. Bello, "Look, listen, and learn more: Design choices for deep audio embeddings," Proc. 2019 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2019, pp.3852–3856, 2019.

[22] High-Level Expert Group on Artificial Intelligence, "Ethics guidelines for trustworthy AI," report, European Commission, 2019.

[23] J.M. Wing, "Trustworthy AI," Communications of the ACM, vol.64, no.10, pp.64–71, 2021.

[24] H. Liu, Y. Wang, W. Fan, X. Liu, Y. Li, S. Jain, Y. Liu, A. Jain,

and J. Tang, "Trustworthy AI: A computational perspective," ACM Transactions on Intelligent Systems and Technology, vol.14, no.1, pp.1–59, 2022.

[25] M. Goto and R.B. Dannenberg, "Music interfaces based on automatic music signal analysis: New ways to create and listen to music,' IEEE Signal Processing Magazine, vol.36, no.1, pp.74–81, 2019.

[26] P. Knees, M. Schedl, and M. Goto, "Intelligent user interfaces for music discovery," Transactions of the International Society for Music Information Retrieval, vol.3, no.1, pp.165–179, 2020.

[27] "Group session - Spotify." https://support.spotify.com/us/article/group-session/.

[28] "JQBX - Listen Together. DJ Online. Discover New Music." https://www.jqbx.fm/.

[29] K. Yoshii, M. Goto, K. Komatani, T. Ogata, and H.G. Okuno, "Hybrid collaborative and content-based music recommendation using probabilistic model with latent user preferences," Proc. 7th International Conference on Music Information Retrieval, ISMIR 2006, pp.296–301, 2006.

[30] M. Tiemann, S. Pauws, and F. Vignoli, "Ensemble learning for hybrid music recommendation," Proc. 8th International Conference on Music Information Retrieval, ISMIR 2007, pp.179–180, 2007.

[31] K. Yoshii and M. Goto, "Continuous pLSI and smoothing techniques for hybrid music recommendation," Proc. 10th International Society for Music Information Retrieval Conference, ISMIR 2009, pp.339–344, 2009.

[32] Z. Xing, X. Wang, and Y. Wang, "Enhancing collaborative filtering music recommendation by balancing exploration and exploitation," Proc. 15th International Society for Music Information Retrieval Conference, ISMIR 2014, pp.445–450, 2014.

[33] A. Vall, M. Skowron, P. Knees, and M. Schedl, "Improving music recommendations with a weighted factorization of the tagging activity," Proc. 16th International Society for Music Information Retrieval Conference, ISMIR 2015, pp.65–71, 2015.

[34] D. Liang, M. Zhan, and D.P.W. Ellis, "Content-aware collaborative music recommendation using pre-trained neural networks," Proc. 16th International Society for Music Information Retrieval Conference, ISMIR 2015, pp.295–301, 2015.

[35] R.S. Oliveira, C. Nóbrega, L.B. Marinho, and N. Andrade, "A multiobjective music recommendation approach for aspect-based diversification," Proc. 18th International Society for Music Information Retrieval Conference, ISMIR 2017, pp.414–420, 2017.

[36] O. Gouvert, T. Oberlin, and C. Févotte, "Matrix co-factorization for cold-start recommendation," Proc. 19th International Society for Music Information Retrieval Conference, ISMIR 2018, pp.792–798, 2018.

[37] L. Baltrunas, T. Makcinskas, and F. Ricci, "Group recommendations with rank aggregation and collaborative filtering," Proc. 4th ACM Conference on Recommender Systems, RecSys 2010, pp.119–126, 2010.

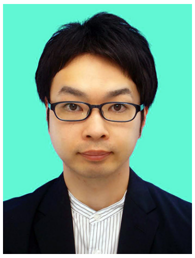
[38] S. Qi, N. Mamoulis, E. Pitoura, and P. Tsaparas, "Recommending packages to groups," Proc. IEEE 16th International Conference on Data Mining, ICDM 2016, pp.449–458, 2016.

[39] D. Serbos, S. Qi, N. Mamoulis, E. Pitoura, and P. Tsaparas, "Fairness in package-to-group recommendations," Proc. 26th International Conference on World Wide Web, WWW 2017, pp.371–379, 2017.

[40] L. Xiao, Z. Min, Z. Yongfeng, G. Zhaoquan, L. Yiqun, and M. Shaoping, "Fairness-aware group recommendation with pareto-efficiency," Proc. 11th ACM Conference on Recommender Systems, RecSys 2017, pp.107–115, 2017.

[41] D. Sacharidis, "Top-N group recommendations with fairness," Proc. 34th ACM/SIGAPP Symposium on Applied Computing, SAC 2019, pp.1663–1670, 2019.

[42] M. Stratigi, J. Nummenmaa, E. Pitoura, and K. Stefanidis, "Fair sequential group recommendations," Proc. 35th Annual ACM Symposium on Applied Computing, SAC 2020, pp.1443–1452, 2020.



Kosetsu Tsukuda received the Ph.D. degree in Informatics from Kyoto University, Japan in 2014. He is currently a Senior Researcher at the National Institute of Advanced Industrial Science and Technology (AIST), Japan. His research interests lie in the areas of recommender systems, user generated content, and user behavior analysis. He has received 17 awards, including IPSJ Computer Science Research Award for Young Scientists and IPSJ Yamashita SIG Research Award.

Keisuke Ishida was the CEO of Ohma Inc. and managed an online social network mining system, SPYSEE, until 2011. Since 2012, he has been a Technical Staff (Creative Engineers) at the National Institute of Advanced Industrial Science and Technology (AIST), Japan. He has implemented web applications regarding musical information processing and information visualization.

Masahiro Hamasaki received the Ph.D. degree in informatics from the Graduate University for Advanced Studies (SOKENDAI), Japan in 2005. He is currently a Senior Planning Manager at the National Institute of Advanced Industrial Science and Technology (AIST), Japan. His research interests include Web mining, semantic Web, and social media analysis. He is a member of the JSAI, the IPSJ, and ACM.

Masataka Goto received the Doctor of Engineering degree from Waseda University in 1998. He is currently a Prime Senior Researcher at the National Institute of Advanced Industrial Science and Technology (AIST), Japan. Over the past 31 years he has published more than 300 papers in refereed journals and international conference proceedings and has received 65 awards, including several best paper awards, best presentation awards, the Tenth Japan Academy Medal, and the Tenth JSPS PRIZE. He has served as a committee member of over 120 scientific societies and conferences, including as the General Chair of ISMIR 2009 and 2014. As the research director, he began the OngaACCEL project in 2016 and the RecMus project in 2021, which are five-year JST-funded research projects (ACCEL and CREST) related to music technologies.