



HAL
open science

Karaoke Entertainment Character Based on User Behavior Recognition

Minkyu Choi, Junichi Hoshino

► **To cite this version:**

Minkyu Choi, Junichi Hoshino. Karaoke Entertainment Character Based on User Behavior Recognition. 16th International Conference on Entertainment Computing (ICEC), Sep 2017, Tsukuba City, Japan. pp.431-434, 10.1007/978-3-319-66715-7_56 . hal-01771287

HAL Id: hal-01771287

<https://inria.hal.science/hal-01771287>

Submitted on 19 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Karaoke Entertainment Character Based on User Behavior Recognition

Minkyu Choi¹ Junichi Hoshino²✉

¹ University of Tsukuba, Graduate School of Systems and Information Engineering,
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577 Japan

² University of Tsukuba, Systems and Information Engineering,
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573 Japan

Choi.minkyu@entcomp.esys.tsukuba.ac.jp,
jhoshino@esys.tsukuba.ac.jp

Abstract. In recent years, research on CG technology has advanced. However, the current CG character can only perform deterministic behavior; there is a problem that interaction cannot be real. Therefore, it is necessary to change the facial expression and gaze in real time to the user by the CG character to select a specific user in the space where there are multiple users, and to do real interaction with the user by using the episode control system and user recognition. In this paper, Kinect V2 recognizes the behavior of a specific user from multiple users, generates a complex gesture of CG characters, gaze representation etc., and proposes a system that CG character supports the provision of services. In this paper, we verify behavior recognition when there is one user and verify behavior recognition by priority when there are multiple users. Because there were many misrecognition of the skeleton, there were many times when the behavior was not recognized. In the evaluation experiments, we will realize even CG character's behavior generation in relation to episode control technology.

Keywords: multi-user recognition • behavior generation • CG character.

1 Introduction

Karaoke is an entertainment that only the musical accompaniment of each song is recorded, and users sing along with the accompaniment when it's played. In Japan, 'karaoke boxes' that provide karaoke became popular in 1980's, and ever since, they have been widely used for friends' gatherings, social occasions, and 'after-drink parties' as well as stress release purposes. In 2015, there are over 150,000 karaoke places and the market size of the karaoke boxes is said to be about 400 million yen. Karaoke is also widely popular in Asia, America, and in Europe, and there is the World Grand Prix of karaoke held as well. It is also impacting the music industry as popularities of songs in Karaoke influences the sales of CDs and downloads as well.

A typical karaoke room in Japan normally has a large monitor that displays promotion videos or image videos of songs and devices to select songs. As users input songs

they would like to sing, it will play them in turn. Sometimes, it has scoring system to judge how good they sang their songs, or video recording system that records them singing and post it on SNS. However, there is a room for improvement, in terms of some gismos to entertain the user, or pick up his/her preferences and make recommendations of songs and services.

This article proposes an entertainment character that, when multiple users are using at a karaoke room, recognizes the behaviors of the users to entertain them accordingly, and promote songs and food & drink services to them.

2 Entertainment Character for Karaoke Room

Fig. 1 shows a scene in a karaoke room. A CG character as a companion is displayed on the large monitor. Also, on the table is a small monitor for selecting songs. The character talks with voice to the users, and the users are to select answers displayed on the monitor. During the conversations, the character picks up your preferences, promotes songs and also recommends some foods and drinks as well as other services.

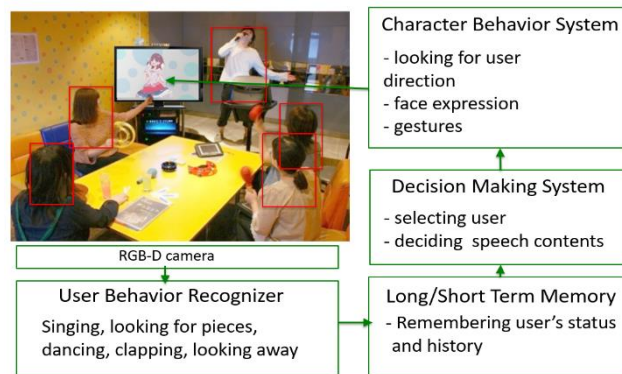


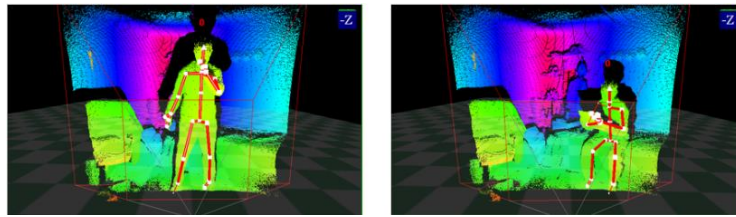
Fig. 1. Scene in a karaoke room and system architecture

In this system, the character recognizes users' behaviors such as singing, dancing and selecting songs, and it would sing along together, help you select songs and such. In this way, the character can avoid doing improper things such as talking to users when they are singing, and it can talk to you in appropriate timing. The judging of the character is done according to the priority degree shown in Table 1, using the results of state recognition on users, by selecting action module, in which some behavior or contents of talks are recorded. When someone is looking sideways, and is not paying attention to the singer, the character talks to motivate him/her to participate. Or, when users are enjoying karaoke, the character would also join them and take behaviors to entertain them such as singing, or clapping its hands. When no one is singing and they are selecting songs, it would recommend some songs, talk about itself, chat about random things based on the news of the day or it would recommend other services.

Table 1. User Behavior and priority

User's state	Acting expression	Priority
(a)Not paying attention to the singer	Focusing one's attention	1
(b)Singing	Livening up the mood, Dancing together	2
(c)Dancing	Dancing together	3
(d)Clapping one's hands	Clapping together	4
(e)Choosing a song	Recommending a song	5

As a method to recognize users's behaviors, as there are many irregular lights such varying lighting and light from monitors, it uses RGB-D camera that can obtain images of certain distances to record users, and obtain their framework information such as bodies and limbs (Fig.2). For extraction of framework information, the kinect library was used. By matching with preliminarily recorded postures and behavior patterns of users, it recognizes users' behaviors. Fig. 2 shows an example of extracting framework information. As shown here, it is capable of detecting basic postures of users in karaoke scenes.

**Fig. 2.** RGB-D image

3 Preliminary Evaluation

In this chapter, how the system recognizes users' behaviors and how the judgment of the character is made that were mentioned in chapter 2 were verified by an experiment.

1) Recognition of single user's behavior and the behavior generation by the character

It was verified, when a behavior of a user is recognized, whether the character was able to generate an anticipated behavior. Table 2 shows the rate of correct recognition when one behavior was showed 10 times. It was able to recognize behaviors from (A) to (D) mostly. However, the recognition rate was down for (E) choosing a song, as the monitor for selecting songs blocked the user's lower body, making it difficult to judge whether the user was sitting or standing.

2) Recognition of two users' behaviors and behavior generation by the character

It was verified whether the character could choose appropriate behavior according to the priority degree of the behavior when there are two users. For instance, each of

the two users took behaviors of (A) and (B), (A) is to be selected as its priority is higher. Table 3 shows, when one pattern of behavior was showed 10 times, the anticipated behaviors of the character and actual behaviors it generated. In many cases, anticipated behavior was generated, but when two users overlapped each other and part of their bodies were blocked, recognition errors tended to occur. For instance, when a user was not paying attention to the singer as in (A), if part of her body was not visible, it was recognized that the user was merely stand-by state.

Table 2. Accuracy of user recognition

User's state	Accuracy of recognition
(a)Not paying attention to the singer	100%
(b)Singing	80%
(c)Dancing	70%
(d)Clapping one's hands	80%
(e)Choosing a song	40%

Table 3. Recognition result of behavior by priority

User 1	User 2	Expected result	Accuracy of recognition
A	B	A	25%
A	C	A	100%
A	D	A	100%
A	E	A	100%
B	C	B	100%
B	D	B	50%
B	E	B	100%
C	D	C	50%
C	E	C	100%
D	E	D	75%

As explained, the recognition capability of RGB-D camera was mostly stable, but recognition errors tended to occur when part of users' bodies were blocked by an object or other users. In such cases, the character would generate unexpected behaviors, and this should be examined in actual karaoke scenes, how disturbing they would be.

4 Conclusion

This article proposed an entertainment character that, in a karaoke room where one or multiple users are singing, recognizes users' behaviors and entertains them in karaoke. For the future projects, the recognition ability and contents of its conversations of the character should be improved by creating a mock karaoke room where users are actually engaged in karaoke to observe whether the reactions of the character are appropriate, and whether it can entertain the users.

References

1. Masao HASE、Hiroshi MORI、Kazuhito SHIRATORI、Junichi HOSHINO : Action control of shop character based on customer state、IPSJ SIG Technical report (EC)2008(129(2008-EC-011))、pp.37-40 (2008)
2. Atsushi NAKANO、Jin KOUMURA、Eriko MIURA、Junichi HOSHINO : Spilant World: Interactive Emergent Story Game using Episode Tree、The Journal of the Society for Art and Science Vol.6(3)、pp.145-153 (2007).