

PARAOKE alpha: a new application development of multiplex-hidden display technique for music entertainment system

Wataru FUJIMURA¹⁾ Yukua KOIDE¹⁾ Takuya SAKAI¹⁾
Songer ROBERT²⁾ Takayuki KOSAKA¹⁾ Akihiko SHIRAI¹⁾

May 17, 2012

1) Kanagawa Institute of Technology, Atsugi, Japan

2) Kanazawa Technical College, Kanazawa, Japan

paraoke (at) shirai.la

Abstract

This project is focusing on the development of a new generation of karaoke entertainment system using augmented reality techniques. It has a multiplex hidden image display using stereoscopic projection and the hiding algorithm “ScritterH”, a special device to see hidden images with a wireless controller called “Fil-Con”, and a dance motion analyzer using Kinect for natural dancing support. The audience can participate in multilingual karaoke by singing and dancing. This system expands the possibility of karaoke entertainment from current karaoke systems meant exclusively for singers.

1 Motivation: Next generation of Karaoke



Figure 1: Karaoke is a singing entertainment system. This project focuses on extending it for wider range of people.

Currently, karaoke, which means “empty orchestra” (in Japanese) is a type of entertainment media used throughout the world. It was invented in 1971, in Japan by Daisuke Inoue who was awarded the “Ig Nobel Peace Prize” in 2004 for his inventing karaoke, thereby providing an entirely new way for people to

learn to tolerate each other [1]. This project attempts to extend karaoke so a wider range of people can enjoy it.

The “PARAOKE” project is focused on solving the exclusiveness of current karaoke entertainment. If the media system can include support for audience members who currently only listen to the singers without interfering, it may be possible to create a new entertainment system that can continuously extend current entertainment culture and user habits. Especially, current karaoke users may be satisfied when there are only a few singers at a time because the systems are designed only for the singer and, in private karaoke rooms, a price is charged to every participant according to the amount of time in some country like Japan. User satisfaction may depend on the amount of singing, not listening.

2 Concept of PARAOKE

PARAOKE is an acronym of “Parallel Augmented Reality for Audience-Oriented Entertainment”.

With modern karaoke systems where the singer is at the heart of the action, any content that would inhibit a song (such as games) cannot be played at

the same time. The display of a karaoke system is an especially important device. Fig.2 shows an isolated singer who is singing in front of an audience that is not looking in the same direction where the song information is being displayed. If a multiplexed display can be applied to a karaoke system, it may be possible to realize a new kind of entertainment that more naturally involves not only karaoke singers but also karaoke audience members.



Figure 2: An issue with current karaoke systems: Most are designed only for the singer, not for the audience, making the singer may feel isolated.

If a multiplexed information display is used for karaoke, it may enhance not only the singer but also dancers and musicians, expanding the enjoyment to be had in arrangement with music.

Fig.3 is a concept image describing the idea behind the PARAOKE (Parallel Augmented Reality for Audience-Oriented Entertainment) system. The singer is on the right but the audience members can freely dance and play guitar. PARAOKE elicits dancing and playing along with the same music from audience members to make the singer feel more comfortable.

Some current music game systems like “Dance Evolution” or “Guitar Hero” motivate players but do not focus on collaboration with the singer. Additionally, conventional music game systems require play in accordance with symbolic scoring within the system. But audience members should not be forced to join the singing by design.



Figure 3: Concept sketch of PARAOKE: Only the singer can see the lyrics through special glasses.

user can see completely different images. ScritterH is an advanced algorithm of Scritter. It has realized free image generation for the naked eye. When users look through a polarization filter, they can see a hidden image. This can enhance current karaoke captions by adding content like music games or menus.

This technology is also compatible with current stereo projectors. Not only for current karaoke systems, Scritter and its series of application can also give new value to the future of stereo 3D imagery.



Figure 4: Multilingual subtitles with Scritter

3 System architecture

3.1 ScritterH: Mutlplex-Hidden Imagery

This project involves the development of a new hiding algorithm called “ScritterH” to realize multiplex-hidden imagery[2, 3, 4, 5, 6, 7, 8]. Scritter is a new application of stereoscopic projectors. It can make multi-channels using polarization filtered glasses. The

3.2 Fil-Con

“Fil-Con”, is a controller with a filter for use with ScritterH. In particular, it consists of a polarization filter and a wireless controller. When using this controller to project multiplexed images from a pair of projectors onto a silver screen, users can see the hidden parallel image. This allows for the configuration of an augmented reality display using optics free of any electronic devices. The prototype (Fig.5) utilizes

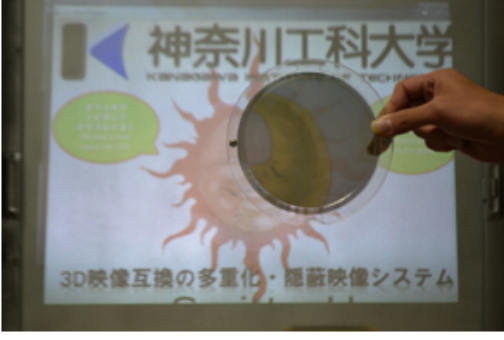


Figure 5: ScritterH: This method can show parallel images without any electronic devices.

a WiiRemote as a wireless controller. It can detect its direction and control an alternate menu system (fig.6). Multiplexed imagery with a stable controller is useful in karaoke environment.



Figure 6: Fil-Con: Polarization filter and controller

3.3 Natural Dance Analyzer

Natural Dance Analyzer (NDA) is a real-time motion capture and motion evaluation system which is based on a Microsoft Kinect sensor and the past research and game development projects as GAMIC algorithm by Fujimura et al. [9, 10] and tools by Misumi et al. [11].

Dancers who join as audience members in karaoke may make swinging motions along with the music. These body movements can be captured by the Kinect sensor as a depth image then processed with the Microsoft Kinect SDK to obtain kinematic bone positions of the target without any need for calibration.

Dance movement can be described as a non-discrete process. The NDA recognition algorithm is based on multiple kinematics evaluation functions which are



Figure 7: Fil-Con shows an alternate menu system for karaoke environment.

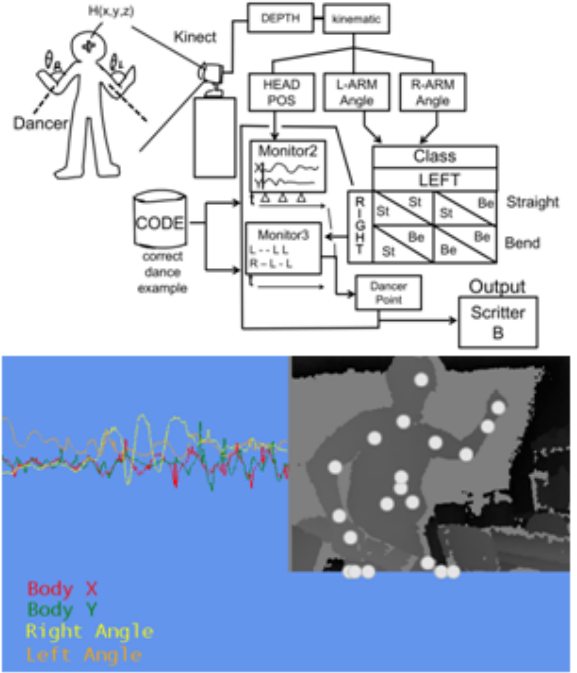


Figure 8: Prototype of Natural Dance Analyzer

configured by dot product between target gesture T_i and the current user's bone posture V_i .

$$f_n = \frac{1}{k} \sum_{i=0}^k \left(\frac{T_i \cdot V_i}{\|T_i \cdot V_i\|} \right) \quad (k : \text{bones}) \quad (1)$$

K is the number of bones, which total at 20 for the entire body in Microsoft Kinect SDK. Each key dance motions have evaluation functions f_0, f_1, \dots, f_n and they evaluate audience players' motion continuously for some characteristic parts like arms. These functions output similarity of the user's posture to a target motion which is defined by ideal or arbitrary values.

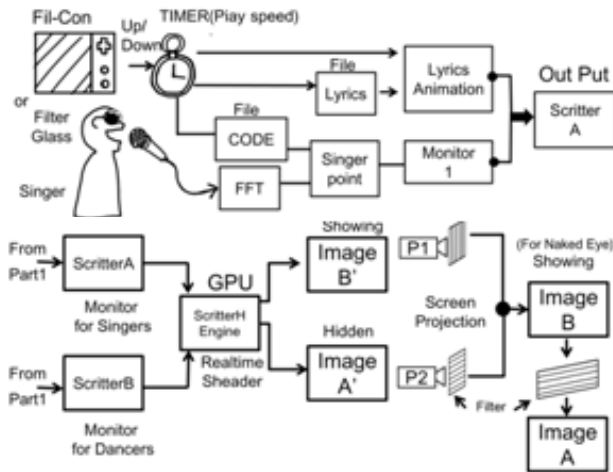


Figure 9: Karaoke evaluation system and ScritterH engine

NDA takes the body position coordinates (x, y) as well as left and right arm angles $(\theta L, \theta R)$ from the full kinematics data set. These are then classified according to four possible conditions such as “straight” or “bend” for each arm. Fig. 8 is a screenshot of the prototype. When the system has target dance data captured as the ideal pose in advance, it evaluates the similarity and rewards points. The point and movement monitors show on the main screen in real-time to further motivate audience activity.

3.4 Karaoke evaluation

To realize the enhancement of current karaoke experiences for an international scene like Laval Virtual ReVolution, a karaoke evaluation system specialized for the ScritterH environment is being developed.

Fig.9 describes a real-time process for singers. The singers who use special glasses or the Fil-Con and microphone can see lyrics with markers (referred to as “Code”) that are generated in advance from the music.

The singer’s voice is processed using a FFT (Fast Fourier Transform) and if it matches the Code, the evaluation function will reward points. The Code and FFT results are shown along with the lyrics in the hidden image to help singers perform foreign or unknown songs more easily. This system also can support karaoke beginners, because the hidden image has an advantage over double screens which are shown to all audience members.

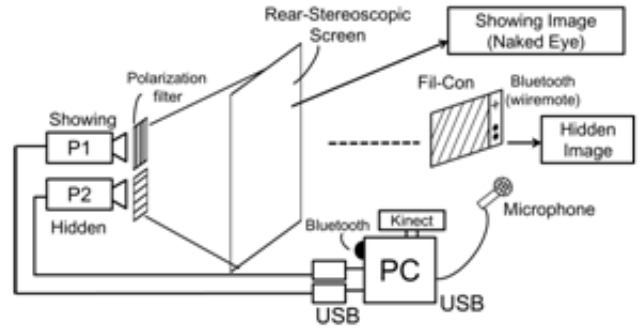


Figure 10: System installation of PARAOKE

4 Trial experiences

Currently, this system is the alpha version of the project concept. Now it is obtaining actual user experiences in natural entertaining conditions. As a field test, it will be exhibited at Laval Virtual 2012. This version allows French people to sing Japanese P-P music with our dance-karaoke system. Audience members viewing the screen without special glasses will see lyrics written in French phonetics. French and Japanese participants both can freely enjoy karaoke together without the language barrier.

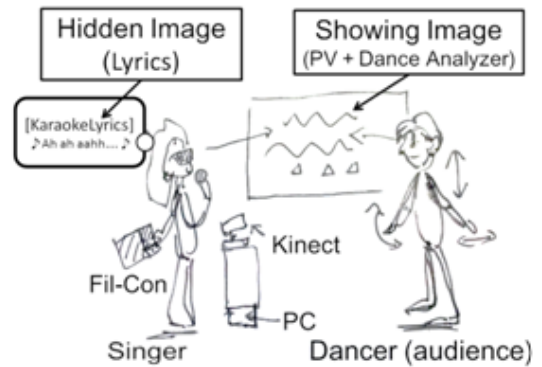


Figure 11: Experience of PARAOKE

ACKNOWLEDGMENTS: We would like to say thanks to Anne FERRERO (Nolife TV) for her internationalization of the lyrics with phonetically corrections.



Figure 12: Beta version of PARAOKE, exhibition in Laval Virtual ReVolution 2012

References

- [1] Stephen Drew. The ig nobel acceptance speeches. *Annals of Improbable Research*, Vol. 10, No. 6, 2004.
- [2] Masao. K. Michimi. I., Mie. S. and Naoki. H. High dynamic range with multiple projectors. *ITE Technical Report, vol.35, no.8*, pp. pp. 45–48, 2011.
- [3] Ian E. McDowall, Mark T. Bolas, Perry Hoberman, and Scott S. Fisher. Snared illumination. In *ACM SIGGRAPH 2004 Emerging technologies*, SIGGRAPH '04, pp. 24–, New York, NY, USA, 2004. ACM.
- [4] Pranav Mistry. Thirdeye: a technique that enables multiple viewers to see different content on a single display screen. In *ACM SIGGRAPH ASIA 2009 Posters*, SIGGRAPH ASIA '09, pp. 29:1–29:1, New York, NY, USA, 2009. ACM.
- [5] Koki N. Takeru U. Takeo, H. and S. Akihiko. Scritter: A multiplexed image system for a public screen. *Proceedings of Virtual Reality International Conference Laval Virtual ReVolution 2010*, pp. pp. 321–323, 2010.
- [6] Oliver Bimber and Daisuke Iwai. Superimposing dynamic range. *ACM Trans. Graph.*, Vol. 27, pp. 150:1–150:8, December 2008.
- [7] Cinema subtitle glasses give promise to deaf film fans. *BBC NEWS TECHNOLOGY*, pp. <http://www.bbc.co.uk/news/technology-14654339>, 2011.
- [8] Koki Nagano, Takeru Utsugi, Kazuhisa Yanaka, Akihiko Shirai, and Masayuki Nakajima. Scritterhdr: multiplex-hidden imaging on high dynamic range projection. In *SIGGRAPH Asia 2011 Posters*, SA '11, pp. 52:1–52:1, New York, NY, USA, 2011. ACM.
- [9] SHIRAI Akihiko FUJIMURA Wataru, IWDATE Shoto. Cartoonect: Sensory motor playing system using cartoon actions. In *Virtual Reality International Conference (VRIC 2011)*, pp. pp.27–30, 2011.
- [10] KOSAKA Takayuki HATTORI Motofumi SHIRAI Akihiko FUJIMURA Wataru, MISUMI Hajime. Development of serious game which use full body interaction and accumulated motion. In *NICOGRAPH International 2011*, p. P13, 2011.
- [11] Hajime Misumi, Wataru Fujimura, Takayuki Kosaka, Motofumi Hattori, and Akihiko Shirai. Gamic: exaggerated real time character animation control method for full-body gesture interaction systems. In *ACM SIGGRAPH 2011 Posters, SIGGRAPH '11*, No. 5, pp. pp. 5:1–5:1, Vancouver, British Columbia, Canada, 2011.