SpeechProtector: A Speech Protection System for Preventing Reporting Bias

Kazutaka Kurihara
National Institute of Advanced Industrial Science and Technology
Chuo Dai 2, 1-1-1 Umezono, Tsukuba, Ibaraki, Japan
k-kurihra@aist.go.jp

Koji Tsukada
JST PRESTO
2-1-1 Otsuka Bunkyo-ku, Tokyo, Japan
tsuka@acm.org

ABSTRACT
Freedom of expression is welcomed in democratic nations, but there is no end of cases in which recorded video is processed to report information not intended by the person in the video. For this article, we have developed a prototype system for preventing this sort of bias in reporting. The system is a smartphone application that allows users, who are the subject of news-gathering, to also record the material themselves, post it to a video sharing site, and to display a QR code containing a link to the video. The system enables a link to a video reproducing the original statements to be forcefully embedded in the report video, which should inhibit bias in the reporting as it is presented later. In this article, we discuss development of this prototype system in detail and discuss technical limitations and its significance in society.

Author Keywords
Reporting bias, QR code, smartphone, Youtube.

ACM Classification Keywords
H5.1. Information interfaces and presentation (e.g., HCI): Multimedia Information Systems; H5.2. Information interface and presentation (e.g., HCI): User Interface.

General Terms
Design, Algorithms, Human Factors.

1. INTRODUCTION

According to the Japanese Wikipedia entry for "Journalism"[4], journalism is a type of expression that uses the media. Journalism and the right to knowledge are based on the right to freedom of expression. On the other hand, journalism must also maintain the principle of objective reporting.

In spite of this, in actual reporting, the facts are very often not discussed as-is, but are reconstructed, incorporating the subjective biases of the reporting organization. For recordings of people making statements, the content is often split up or rearranged, contrary to the intentions of the speaker, and this can cause controversy[6].

Figure 1. Receiving materials using SpeechProtector

The remedy to this is generally said to be that the receiver of the information must acquire media literacy and cultivate the ability to discern the truth of information they receive[4]. However, this measure protects the interests of the one receiving the information, not the intentions of the person being reported on or their credibility in society.

The goal of this research has been to reduce bias in journalistic reports that use statements made by people in presentations, speeches, press conferences or street interviews, even if the statements have been manipulated. This is done through technology that enables receivers of the information to access the speaker's statements as originally intended. This protects the credibility of the speaker in society, and can be expected to tighten global monitoring of bias in reporting.

CHI 2012, May 5-10, 2012, Austin, TX, USA.

As the first step in this article, we describe development of the smartphone-based SpeechProtector prototype system for preserving reported statements (Figure 1). The system applies particularly to reports involving photographs and video. The user, who is the subject of the report, also records the entire session, posts it to a video sharing site, and forcefully inserts a link in the material used for reporting by displaying a QR code with a time-stamped link to this video during the whole session. The system provides a simple interface for accomplishing this.

Then, in whatever way the gathered video data and statements are modified, the receiver can easily access the original content using a mobile terminal such as an ordinary mobile phone, as long as the QR code remains in the report. If the user is not satisfied with the report, he/she can appeal to society using this link, through SNS or other personal publishing avenues.

This article is structured as follows. We first summarize related work and then discuss the design of SpeechProtector and implementation of a prototype. Then, we discuss testing the basic performance of the prototype, as well as technical limitations and its significance in society.

2. RELATED WORK

This research is related to protecting the legal right to use of one's likeness. This right includes the right not to have a photograph taken without authorization (right to refuse photograph), and the right not to have such an image used by others (right to refuse use)[14]. However, in Japan, the constitution recognizes freedom of expression, and the right to protect one's likeness is not stipulated, so it is often superseded by freedom of expression[11]. In practice, except for certain people such as performers, whose likenesses have added value, it is difficult for ordinary citizens to refuse reporting if a journalistic organization determines that there is "social need". In this research, we propose a technical base that has the ability to inhibit bias in reporting, without refusing reporting entirely.

This research can also be considered as an application of a type of electronic watermarking. Technologies like [10], from Mitsubishi Electric, are the latest technologies able to embed and detect copyright information in Hi-Vision video. They are mainly useful when copyright holders want to control distribution of information. In this research, however, the person being recorded does not have the right to edit the material collected, so a way to force insertion of watermark information into the report is needed.

We next describe how this additional data can be embedded visually in the real world. For promotional purposes, famous people often place product names or names of organizations to which they belong in the background when they are the subject of news gathering, and this is often seen on television. Learning from this practice, we also aim to have the information included automatically when the user's face is photographed during news gathering, such that it would require a certain amount of work or the image would lose value if the code was removed. QR codes[3] are a 2D-barcode technology developed by Denso Inc. (currently Denso Wave Corp.) to overcome some shortcomings of conventional barcodes, with an explicit statement that patent rights will not be exercised. Today, many mobile devices include a function able to read them, and they have spread to the extent that almost everyone with a mobile phone can read them. They are widely used, such as in paper and various urban advertising media, and on packaging to indicate product qualities. A major way in which they are used is to apply a code to physical objects or content, linking to more detailed information. The particular visual characteristics of QR codes make them easily recognizable, which is another characteristic. Other research on 2D barcodes includes C-Bands, due to Miyaoku et al[9]. C-Bands are a 2D barcode with a ring shape, extending to allow for more-flexible design, but they have not achieved the level of penetration of QR codes, and are not as recognizable.

Embedding codes in the real world has also become important in the field of augmented reality (AR), providing a means of obtaining contextual information. NaviCam[12], proposed by Rekimoto, overlays additional information about an item when it is being photographed, using codes embedded in the environment. Also from Rekimoto, CyberCode[13] enables the 3D coordinates of an object being photographed to be obtained using 2D barcodes. These techniques enabled interaction in a space with a virtual world overlaid on physical objects, and had a significant effect on later AR research. ARToolKit[5] is a toolkit making it easier to build this sort of application, and has been used widely in fields from education and research to hobbies.

3. DESIGN AND IMPLEMENTATION OF SPEECHPROTECTOR

3.1 Problem Analysis and System Design

Behind the problem that reporting can be done contrary to the intentions of the person subject to reporting are facts that it is difficult to refuse information gathering, as discussed in Section 2, and there is no control over how the material gathered will eventually be edited and reported. Thus, we introduce a mechanism that allows the person to force his/her intentions to be reflected in the information gathered, and allows the viewer of the report to notice and access this information. The mechanism involves designing three aspects: (1) how can the person force his/her intentions to be reflected in the report, (2) how can the viewer's attention be drawn to them, and (3) how can they be expressed to the viewer?

For (1), sudden requests for news gathering can be accommodated using a system that operates on a smartphone, which can be expected to be carried at all times. The person being photographed can force the 2D

barcode to be included in the image of his/her face by holding up an image of the barcode, (Figure 1). This action falls under his/her freedom of expression, and cannot be forbidden by anyone else. As long as the resolution is high-enough, this code will still be valid even if sections of the captured video are edited. Similarly, the code should still be valid when images are posted as still photos on the Net or in paper media.

For (2), QR codes can be used for the 2D barcodes in (1). QR codes have a particular visual form and are used widely in the public, so viewers recognize them easily. They can also be read easily using any mobile device in their possession, conveying the original intentions of the person being photographed.

Finally, (3) can be implemented by configuring the destination of a link in the QR code to a video of the original content, captured by the person themselves. This is because the originally stated content is the most effective and impartial way to critique bias in reporting. Further, adding SNS and other functions that enable third parties to discuss the critique with the associated video creates a mechanism that will promote interest in reporting bias within society.

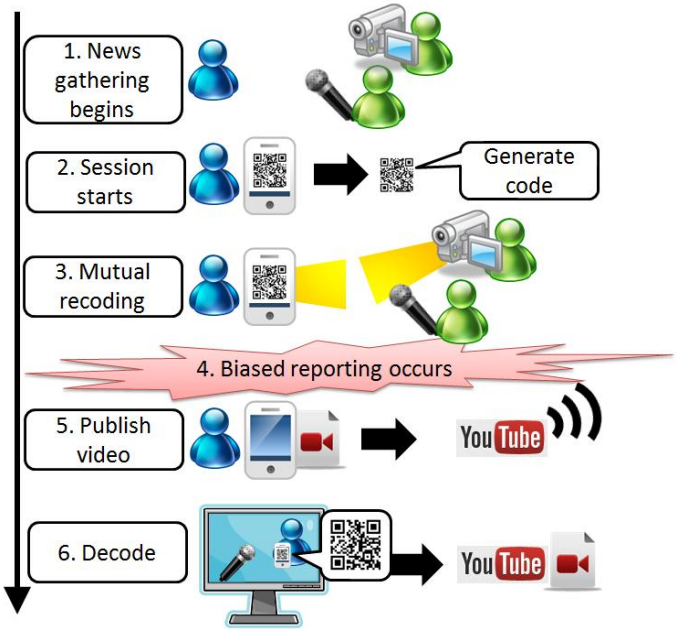

Figure 2. Scenario using SpeechProtector

3.2 Implementation

We implemented the prototype SpeechProtector system for inhibiting bias in reporting using a client-server system. The server uses ASP.NET, and the client is implemented as a Windows Phone (smartphone) application. We used YouTube as the base for video hosting and SNS functions.

3.2.1 Usage Scenario

We outline the operation sequence with a hypothetical scenario (Figure 2) below.

(1)The user is conducting his everyday activities, carrying a smartphone, when confronted by a reporter.

(2)The user starts SpeechProtector on the smartphone, initiating a new session. This generates a QR code for the session.

(3)The reporter begins gathering material (filming). The user holds the smartphone near his/her own face while also recording video with SpeechProtector. While recording, the QR code created in (2) is displayed prominently on the smartphone, and the reporter cannot avoid photographing it when filming the user's face.

(4)Later, if the interview is reported according to the user's intentions, there is nothing to do. In such a case, a message indicating that "no corresponding video has been published" is displayed when the QR code is decoded. If not reported according to the user's intentions, he/she has the following means of opposing it.

(5)Simple operations in SpeechProtector will publish the video taken with the smartphone while the reporter was filming on YouTube. When this is done, the destination of the QR code in (2) is automatically configured to link to the video shared on YouTube. The user can then build up criticism of the biased report using comments on YouTube or other venues such as SNS or Internet bulletin boards.

(6)Later, any ordinary viewer can read the QR code embedded in the biased report using his/her mobile phone, and play the YouTube video of the recorded original statements. Thus the authenticity of the biased report is ascertained and the YouTube comment feature can be used as a starting point for criticism or participation in related discussion.

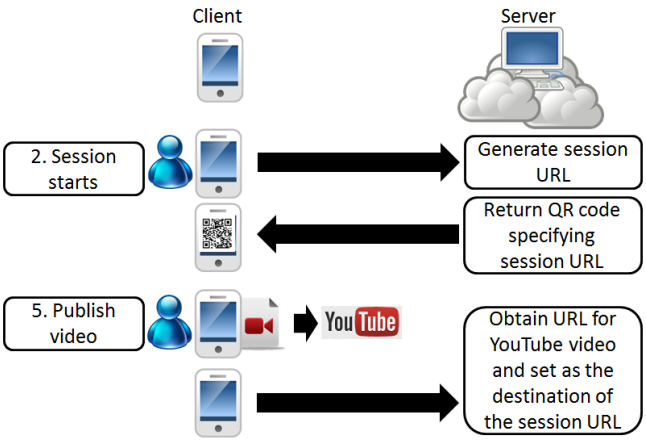

Figure 3. Client-server communication

3.2.2 Client-Server Communication

Communication between client and server is as shown in Figure 3. When the server receives notification to start a session from the client, as in 3.2.1(2), it generates a session URL, which is unique for each session. It then returns a QR code to the client indicating this URL.

The URL for the YouTube video is sent to the server after the video is published on YouTube in 3.2.1(5), and it is configured as the forwarding address of the session URL.

3.3 Progress Bar Preventing Video Cut-and-Paste

During video capture in 3.2.1(3), the QR code is displayed on the user's smartphone screen, together with a full-screen progress bar in the background, which updating regularly every few seconds. This helps prevent editing to bias a report such as cutting out sections, for example, to the statement "I am opposed to AAA, but support BBB" into "I am opposed to…BBB". In such a case, cutting and pasting is likely to cause the progress bar in the video to appear unnatural, making the edit easy to detect.

3.4 Embedding Time Information in QR Codes

During video recording in 3.2.1(3), the timestamp since the start of recording is encoded in the QR code displayed on the smartphone screen, and is updated in intervals of a configurable number of seconds. When the code is decoded, video on YouTube is played starting at the corresponding time. This enables an ordinary user to access the corresponding video by decoding the QR code easily, for comparison with the video in the media report.

4. TESTING READABILITY OF QR CODES

In this section we discuss a basic evaluation of appropriate sizes and shooting angles for the QR codes presented by SpeechProtector. The important aspect for this system is that the QR codes displayed in section 3.2.1(3) be readable by mobile terminals carried by ordinary viewers in 3.2.1(6). This involves factors that the user can control and others that he/she cannot.

Factors that the user cannot control include the resolution that the reporter uses for capture and presentation, and of the presentation equipment—whether TV, PC screen or paper media—that viewers see. In most cases today, the resolution is quite high in both cases, and often HD quality (1920 x 1080 pixels). The size of the QR code displayed on the presentation equipment being viewed must also be considered, but this does not have much effect on readability because viewers are able to zoom in on the code and place the code reader as close as possible to optimize reading.

Factors that the user can control include the relative size of the QR code within the field of view of the camera. We adopted a strategy of displaying the QR code with size at least as big as the main features of a person's face. Specifically, we display a QR code on the smartphone with sides of approximately 6 cm. This is approximately the same as the average interpupillary distance for young male and female adults (6.07 cm) according to the AIST Human measurement and form database[1]. When the main objective of a photograph is the face, the face occupies the major part of the field of view, so if the user displays the smartphone near his/her own face, the QR code, which is at about the same scale, is likely to be readable. For example, Figure 1 shows a single frame clipped from a video taken with full-HD quality. The area occupied by this QR code is 114 x 125 pixels, and would be a 1 cm square if output on an A4 sheet of paper. This would of course be readable when printed on paper, which is usually 300 or a higher dpi, and also on a PC display, which is usually approximately 96 dpi.

The QR codes used by the current prototype are version 3, with level Q correction. This configuration of QR code is 29 x 29 cells (basic pattern units). If a single cell is displayed with one pixel, the minimum required to preserve the code is with eight-cell margins horizontally and vertically, giving 37 x 37 pixels. This gives a square of approximately 1 cm when displayed at 96 dpi on a PC display. Thus, a size of 1-cm-square or greater is one criteria for readability when displaying QR codes of this configuration at actual size on a PC screen, ignoring error correction.

We next examine the relationship between the angle at which the QR code is photographed and readability. We displayed a QR code with approximately 6 cm edges on a SpeechProtector terminal and photographed it with a sufficiently-high-resolution digital camera (3646 x 2736 pixels) at a distance of 30 cm and at angles in five-degree intervals from zero to 90 degrees (the angle from the normal to the code surface). These images were displayed at actual size on a PC screen to be read. By using a QR code image with adequate resolution, this configuration enabled investigation of the upper limits for readability, dependent only on the angle. We used the code reader built into the Windows Phone, the iPhone "Renzoku QR Code Reader" applet, and the Android "QuickMark" applet to read the QR codes. When reading the codes, we allowed the distance from the QR code to the reader terminal to be adjusted the freely.

Simply stated, the results depended on the decoder algorithms used in each decoder, but all terminals were able to read the codes at angles from zero to 75 degrees. This suggests that as long as the resolution is high enough, the codes are readable at a wide range of angles, and adequate for practical application.

5. DISCUSSION

In this section, we discuss the impact that SpeechProtector could have on society as well as its technical limitations.

5.1 inhibiting Effect

When a user involved in news gathering displays a QR code, which is a foreign object, near his/her face, it expresses indirectly to the news gatherer that if he/she creates an unfair report, the user is prepared to reveal the truth. This should be effective as social activism to prevent the gathering of materials that lend themselves to such behavior and also misuse of information gathered. Simply from its appearance, the QR code is either meaningless or an intrusion for the news gatherer. It only detracts from the aesthetics, so it is expected to prompt reporters to hesitate or avoid photographing it. We next consider whether news gatherers would manipulate the images of QR codes in gathered materials. Unfortunately, with the current implementation, it is technically possible to suppress or manipulate the QR codes easily using image processing. However, this sort of image processing would decrease the value of the face image, if even just a little, and the unnatural video output would give the impression that something was strange. If the system became widely known, and reporters acquired a tendency to watch for people displaying QR codes when gathering materials, this research could be said to be functioning effectively.

Note that the proposed method is not effective for reporting with pure text, radio or other sound-only media, and not using media in the image modality, such as photographs or video. Shimizu et al.[8] have proposed a mechanism allowing large numbers of listeners in an audience to experience a live performance linked to their smartphones by incorporating additional information into sound with a high threshold of hearing. It is conceivable to incorporate this sort of audio watermarking technology into reports using only sound, but points such as whether ordinary users would even be aware that the watermark was present have room for discussion.

5.2 Reciprocal Monitoring Effects

As mentioned earlier, it is technically possible to conceal or manipulate QR codes easily, but our method would still be effective for situations such as a press conference or lecture, where we can expect that many reporting organizations gather, unless all reporting organizations agreed as a whole to conceal or manipulate the QR codes. This is because if even one of the organizations reported the QR codes truthfully, that organization would gain the right to critique other organizations that are not reporting honestly. This is a reciprocal monitoring effect for reporting bias in reporting organizations. If, hypothetically, all news reporting organizations began reporting this way, it would cause great mistrust of the entire industry when it came to light, causing increased scrutiny for reporting bias in the world, and achieving the objective of this research in a different form.

5.3 QR Code Duplication

The important aspect of this system is to forcefully introduce the QR code into the reporter's video in a form that it will be readable. To achieve this, the QR code should be large, and multiple QR code should be shown. QR codes are easy to duplicate or print, whether on paper or electronic media, so after completing 3.2.1(2), the user can use a spare tablet or PC to repeatedly duplicate the QR code, or print it and fill up the reporting scene (Figure 4, 5). When doing this, the SpeechProtector terminal can be placed on the podium to record the speaker's statements. This is particularly effective when the user can prepare for the news gathering beforehand, such as for a press conference or lecture.

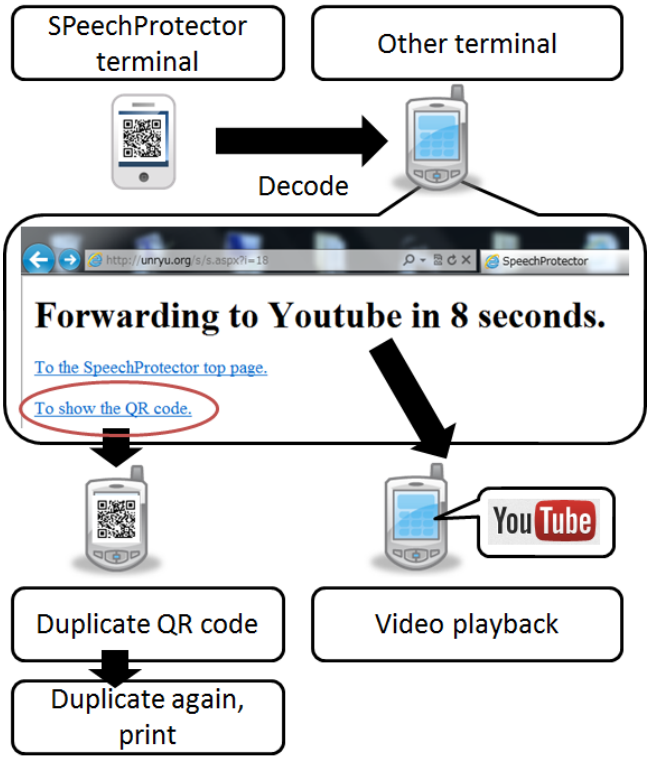

Figure 4. QR code duplication

Figure 5. Example of presenting many QR codes in an interview or lecture

5.4 Audio and Video Recorded by the User

When the user records video using SpeechProtector in 3.2.1(3), the most important factor is to have a record of the original statements by recording the user's voice. On the other hand, the video image can be of either the user or the reporter, depending on where the camera is mounted on the smartphone. If the phone has multiple cameras, the system allows selection of which camera to use.

5.5 Timing the Release of a Video

With the proposed system, we adopt a policy that the user can publish the video he/she has taken himself/herself whenever they become aware of bias in the reporting. This is effective for media that viewers access asynchronously and repeatedly, such as news on the Internet. For television and other broadcast and transient media, publishing after the broadcast would not ordinarily reach viewers. However, news programs broadcast the same content repeatedly, and it is now common to record television programs and watch them at a more convenient time, so this approach should be somewhat effective. Note that the system also allows the user to publish his/her own video before the news material is released. This would avoid the above difficulty, but may be somewhat excessive in moderating the reporter's behavior, so users need to decide according to their own ethical sense.

5.6 Target Users

Users targeted in this research are those concerned about bias in reporting of their own statements. As such, it may be criticized as useful for only a small proportion of the population, such as politicians and celebrities, who often have press conferences. However, opinions of individual people are always being transmitted to the many, suddenly and probabilistically, as in the classical person-in-the-street interview in mass communication, and also with so-called "flare-ups"--the explosive and transient information flows--deriving from recent communications media such as SNS. Providing this system for such situations would be an effective means of managing risk for individuals in a sophisticated information society. Today, smartphones are widespread, and providing this system widely to users on these platforms as a mobile tool for personal protection is a possibility.

5.7 Handling Falsification by Users

One issue with the system is that a video unrelated to that of the reporter could be associated with the QR code. By using this possibility maliciously, users could revise their original statements after the fact and present them as the original. The system implements a minimum level of security to prevent falsification, ensuring genuine communication between client and server and authenticating so that the same client is transmitting continuously.

5.8 Potential for Continuous Streaming

The approach of the proposed system is to use a smartphone, which the user always carries, to record video and publish it when necessary. It has also become practical to relay and publish a constant stream of a user's life using a wearable camera such as the Looxie2[2]. It would also be possible to restructure the proposed system to use such a camera, dynamically incorporating display of QR codes, and enabling particular sections of the constantly recording video to be extracted and published. Such a system is left for future consideration.

5.9 Application for Lectures and Presentations

The proposed system was designed to deter bias in reporting, but it could also be used to enhance the user's expressiveness in lectures or presentations by displaying QR codes near himself/herself or in the presentation slides. Recently, it is not unusual to upload video of the sequence of slides with some photographs for lectures and presentations, and it is technically possible to provide viewers with links to additional information such as the lecture video, promotional video, or the slides.

6. CONCLUSION

In this article, we have proposed the SpeechProtector system to control information bias in reporting, by helping the person being reported on to make his/her own recording of the news gathering, to post it on a video sharing site, and to force a link to this video into the news material being gathered, in the form of a QR code. We implemented a prototype of this system, tested its basic performance, and discussed its technical limitations and significance in society.

ACKNOWLEDGMENTS

A part of this research was instigated by the Japan Science and Technology Agency and received support from a Grant-in-aid for Scientific Research (KAKENHI 23700155).

FUTURE VISON

We are conducting research on new protective measures using information technology, to protect the secure lifestyles of each individual in society. Even when excluding clearly-illegal behaviors such as bodily violence, modern people are subject to various attacks that are not clearly illegal. Some means of protection is needed to oppose such behavior and ultimately to prevent it.

At WISS'10, we proposed a speech inhibiting system called SpeechJammer that uses delayed audio feedback to control speech violating set rules and manners[7]. This and SpeechProtector, proposed in this article, are based on the following four points of focus as characteristics of modern means of self-preservation.

1. Counter-attacks are only permitted if the other party attacks first.

2. The attack from the other party itself is the driving force for a counter-attack, and the counter attack automatically stops if the attack stops.
3. The attack and counter-attack are quantitatively equivalent.
4. Even if the attack is reciprocal, they should not result in destruction.

It has been difficult to apply these sorts of criteria in deciding when and how much to use weapons typical of pre-modern times, but with the development of information technology today, there is hope for solving some aspects of this. Our vision for the future is to pursue such solutions.

With SpeechJammer, it is left to the user of the weapon to decide what is considered an attack, so we are not, in principle, able to prevent misuse of the weapon. On the other hand, with SpeechProtector, we attempted to solve this issue using a mechanism in which the decision depends on society (society can compare the news video with that taken by SpeechProtector and ascertain which is the truth).

However, at base reporting and journalism should have the intention of being entrusted to bring truth to the people and to resist using their influence to manipulate information. It is suggestive that a proposal for preventing bias in reporting ultimately leaves responsibility for the substance of reporting with each individual.

REFERENCES

1.AIST 人体寸法・形状データベース. http://riodb.ibase.aist.go.jp/dhbodydb/index.php.ja.

2. Nexx Wearable Cam Looxie2. http://www.nexx.co.jp/products/life_style/looxcie2/.

3. QR Code.com. http://www.qrcode.com/.

4. Wikipedia「報道」. http://ja.wikipedia.org/wiki/報道.

5. H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In Proceedings of IWAR'99, pp. 85-94, 1999.

6. 荻上 チキ, 飯田 泰之 謙介. ダメ情報の見分け方. NHK 出版, 2010.

7. 栗原 一貴, 塚田 浩二. SpeechJammer:聴覚遅延フィードバックを利用した発話阻害の応用システム. WISS'10 予稿集, pp. 77-82, 2010.

8.清水 基, 平林 真実, ビュトナー クレメンス, 赤松 正行. 高可聴閾音を利用した DTMF 通信によるアーティストと観客のインタラクションの実現. インタラクション 2012 予稿集, pp. 825-830,2012.

9.宮奥 健人, A. Tang, S. Fels. C-Band: 柔軟なカラー環状タグシステム. インタラクション 2006 予稿集, pp. 3-10, 2006.

10.三菱電機. 映像電子透かしソリューション. http://www.mitsubishielectric.co.jp/news/2011/pdf/1027-a.pdf.

11.村上 重俊. 知っておくべき肖像権判例集（講演）. http://www.lawyers-office.net/gyomu/topics/img/topics06.pdf.

12.暦本 純一. 状況認識に基づく新しいインタラクションスタイルの提案. WISS'94 予稿集, pp.9-17, 1994.

13.暦本 純一. 2 次元マトリックスコードを利用した拡張現実感の構成手法. WISS'96 予稿集, pp.199-208, 1996.

14.日本写真家協会会報. 肖像権・撮る側の問題点. http://www.jps.gr.jp/rights/pdf/127_30-31.pdf.