

## Contact Bit:映像上で選択できる光信号の研究

光藤雄一†

本稿では、PointSpeech と呼ばれる音声情報支援システムを提案する。PointSpeech は、カメラで実世界を撮影し、この映像上で希望の場所に指を重ねる、「テレクリック」と呼ばれるアクションを行うことで情報支援を受ける。本研究の特徴は、PointSpeech システムを実装する際に、映像処理による情報取得を避け、「ポインティングデバイス並みの空間分解能を持った光空間通信システム」を作成することでこのような機能を実装したことである。光空間通信システムを応用することにより、「テレクリック」された対象物からユーザに対して、音声情報のような大容量のデータを直接伝送することが可能になり、簡便なシステム構成を実現することができた。

### ContactBit:A Research of Optical Signals Selected with Real World Image

MITSUDO YUICHI

In this paper, we introduce PointSpeech as a real world audio assistant system. This system provides a user to audio assistance about the pointed object by the user. A characteristics of this system is implemented without any image processing, but optical signal processing. By this architecture, pointed object is able to send large amount of data, such as sound file, to user, and this realizes simple bottom-up architecture.

#### 1. はじめに

実世界環境<sup>1)</sup>の研究の一つに、実世界のユーザに対して情報支援する手法の研究がある。この研究でよく取られる方法の一つに、ユーザがあるものに近寄ると情報支援が開始されるシステムがある<sup>2)~4)</sup>。このシステムは、博物館や美術館、イベントなどで運用されている。これらのシステムは接近することで自動的に音声情報支援が開始される仕組みを持つ。

これに対して本研究では、実世界ポインティングデバイスと組み合わせ、「ポイントした情報機器から情報支援が受けられる」PointSpeech と呼ばれるシステムを作成した。

本研究では、情報機器をポイントする際にテレクリックという(図1)手法を使用する。

テレクリックは、映像上で目標に対して指を重ねる仕草である。これは実世界のデバイスを選択する手法の一つであり、デスクトップ環境のクリックを模している。普及しているデスクトップ環境のウィンドウシステムでは、ディスプレイ上で、マウスカーソルをアイコンに重ねることによって情報を選択している。これに対し、テレクリックは、実世界を撮影し、撮影した映像上である情報機器に指などを重ねることによって選択する。この場合、映像がディスプレイ、指がマ

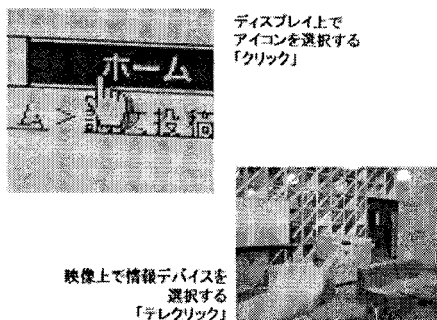


図1 テレクリックの様子

Fig. 1 Tele-Click:Real World Pointing Method

ウスカーソルに対応し、実世界上のデバイスの映像がアイコンに該当する。

テレクリックの特徴は、あくまで映像上で指が重なっていれば良いので、対象物が離れていても選択することが可能なことである。

#### 2. テレクリックと PointSpeech

##### 2.1 テレクリックによる対象物の検知

ユーザがある情報機器をテレクリックした際に、撮影用のカメラと指、情報機器が直線上に並ぶ(図2)。

† 公立はこだて未来大学

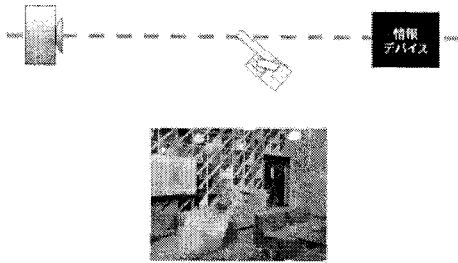


図 2 テレクリックの検出:テレクリック時のカメラ、指、情報機器の位置関係

Fig.2 How to detect tele-click:an geometrical model of tele-clicked information device

本研究では、光空間通信システムを利用し、図 2 の位置関係を満たした場合にのみ光信号を受信することが可能な通信システムを作成した。つまり、このシステムは、厳密な意味でのポインティングデバイスではなく、実世界認識機構なども利用していない。「テレクリックされた送信器からのみ光信号を受信できる」機能を持った情報伝送システムである。

本研究では、この伝送システムから音声情報を送信することで、「テレクリックされた送信器から音声情報を得る」機能を持つ、PointSpeech と呼ばれる機材を作成した。

## 2.2 PointSpeech

PointSpeech は、情報伝送に光信号を利用している。この情報伝送システムは、送信側は 2 つの発光器を備えた光発信器からなり、受信側はカメラ等の撮像素子と光信号を受信するための受光素子から構成される。

受信側の特徴は、撮像素子に向けて受光器が設置されていることである。この構造により、撮像素子から見えない送信器 (= テレクリックされた送信器) から発せられた光信号が受光器から遮断される。しかし、テレクリックしたときに信号が遮断されたのでは都合が悪い。そこで、ユーザがテレクリックしたときのみ光信号が現れるように、送信器に次のような工夫をする。送信器に 2 つの発光素子 (A1 および A2) を装着し、この発光素子から反転信号を出力する (図 3)。

このため、ユーザがテレクリックをしていない場合は、反転した信号が足しあわされて受信器に到達するので、直流成分となり、変調成分である信号成分を検出することができなくなる。ユーザが片方の発光器に対してテレクリックを行った場合にのみ、反転信号の片方が遮断されるため、受信器が信号成分を検出することができる。

## 2.3 システムの概要

構築したシステムは、光信号を受光するためのフォ

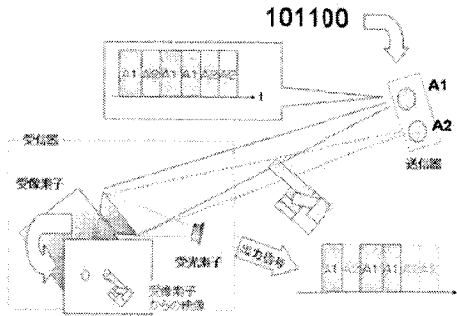


図 3 システムの概要  
Fig.3 Overview of system

トダイオードに代表される受光素子と、CCD に代表される受光素子の 2 つの部分からなる受信器と、一組の LED (発光素子) を備えた送信器から構成される (図 3)。

### 2.3.1 受信器

本研究で使用した受信器は、受光素子として松定プレジジョン製のアバランシェホトダイオードである M5844 を利用し、受像素子としてピンホールカメラを利用した。

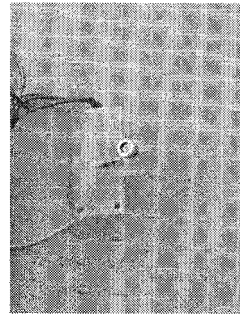


図 4 ホトダイオードと駆動回路  
Fig.4 An image of photo diode

M5844 は単位入射放射束あたり 0.4A の光電流を発生する能力があり<sup>\*</sup>、印加電圧  $V_f$  を加えることでこの光電流を 100 倍まで増幅<sup>\*\*</sup>することができる<sup>\*\*\*</sup>。このホトダイオードの遮断周波数  $f_c$  は 600MHz である (図 4)。

また、ピンホールカメラは F01-3820-IR を使用した。このカメラは、水平画角約  $7.1^\circ$  で  $510 \times 492$  画素の映像を NTSC コンポジットビデオ信号で出力す

<sup>\*</sup> 放射感度  $R=0.4A/W$

<sup>\*\*</sup> 増幅率  $M_1 = 100$

<sup>\*\*\*</sup> 電子アバランシェ現象

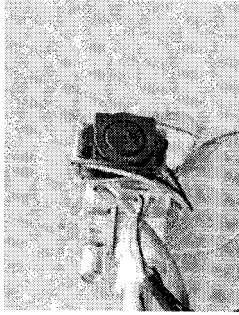


図 5 ピンホールカメラの映像  
Fig. 5 An Image of image sensor

る機能を備える (図 5)。ピンホールカメラの表面は、ピンホール以外の場所は梨地仕上げの黒の樹脂で形成されている。従って、受光素子をこのピンホールカメラに向けて設置した場合、ピンホール内部の光学装置およびピンホールの周囲の樹脂から反射された信号光が受光素子に入射する。本研究では、梨地仕上げの部分で信号光が乱反射し、その一部が受光素子に入射している。

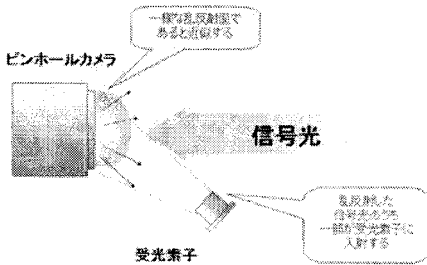


図 6 反射信号光の受信  
Fig. 6 Receiving reflected signal

### 2.3.2 送信器

本研究では、光信号の発信器として 2 対の発光装置を備える光信号発信装置を作成した。これら 2 対の発光装置は、それぞれ 3 個の発光素子である赤外線 LED DNK318U から構成され、6.0cm の間隔で配置されている。これらの発光素子は 50mW (順電流  $I_F=50\text{mA}$  の時) の赤外線放射束を発する能力を持ち、そのピーク発光波長は 870nm、指向半値幅は  $20^\circ$  である (図 7)。また、発光の立ち上がり・立下り時間  $t_r, t_f$  は 7ns であり、高速信号を放射するために十分な性能を持つ。

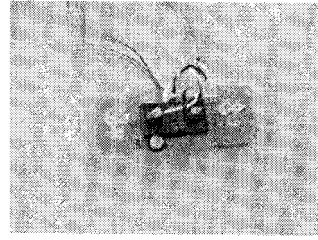


図 7 発光器の映像  
Fig. 7 An Image of transmitter

## 3. テレクリック検知の実証

### 3.1 映像と光信号の整合性

テレクリックを実施するためには、発信器に対してテレクリックが行われた際に、正確に信号が受信されることが最も重要なポイントである。この点を確認するために、映像上のテレクリックの状況と、信号の受信状況を比較する必要がある。そこで、本研究では図 8 のような実験条件の元で光信号の受信を試みた。

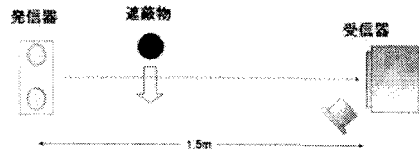


図 8 実験条件  
Fig. 8 Experiment

この実験条件では、カメラの位置と発信器の距離が 1.5m あるため、カメラのそばにいる人物は、指などで発信器に直接接触する事ができず、テレクリックによる選択のみが可能である。この実験は、想定される使用状況を模して蛍光灯点灯下の室内で行われた。信号の受信結果を図 9 に示す。更にカメラから 60cm の

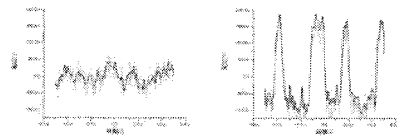


図 9 テレクリックの実証実験  
Fig. 9 Exp:Tele-Click and Optical Signal



距離に直径約 1.5cm の円柱を配置し、これを動かすことによって受光信号がどのように変化するかを測定した。円柱が A 1 の発光素子をテレクリックしている間は光信号が受信されるが、これが移動して A 2 の発光素子をテレクリックすると、信号光が反転するはずである。発光素子 A 1 および A 2 は 6 cm の距離をおいて配置されているため、カメラに対して約 3° の画角を占めている。図 7、10 より、発信器の映像に応じて光信号が反転したことが分かる。従って、3° の分解能で信号光を取得することが可能であることが分かる。以上の実験から、ユーザのテレクリックによって約 3° の分解能で特定の光信号を受信することが可能であることが示され、目標どおり、映像に対応した形でこの光信号が変化することが分かった。

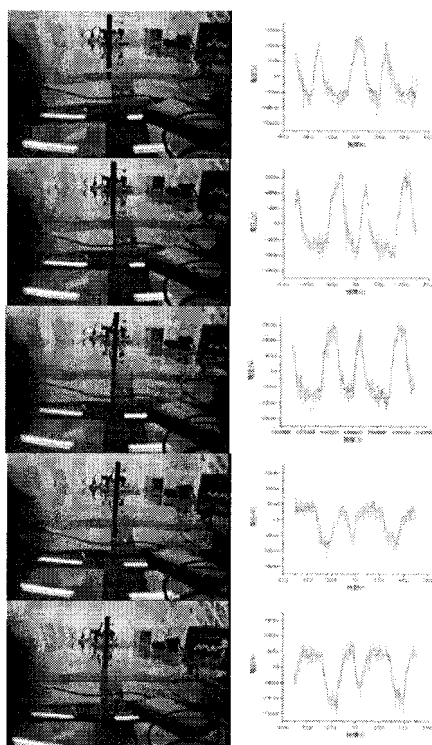


図 10 PointSpeech の空間解像度 (受信シグナル)  
Fig. 10 Experiment 1

#### 4. PointSpeech:光信号による音声情報支援

3 章までの実験により、ユーザのテレクリックによって光信号を検出する仕組みを整えることができ、所定のパルス波形を送信し、受信することが可能であることが実証された。

そこでこの章では、光信号によってどのように音声情報を伝達するかを考える。音声情報を伝送する変調方法として一般的に普及しているものは、

- AM 変調
- FM 変調
- パルス変調

の 3 つである。本研究では、光信号を利用していることと、反転信号を生成する際の簡便性から、パルス変調に注目した。パルス変調は、振幅が一定のパルスによって送信信号を変調する方法である。パルス変調には幾つかの種類があり、音声情報のようなアナログ信号を伝送するための方法として、

- (1) パルス幅変調
- (2) パルス密度変調
- (3) デルタ変調

の 3 つが主に使用されている。パルス幅変調はパルスの幅 (デューティ比) を変化させることで電力を制御し、情報を伝達する。パルス密度変調はパルスの密度を変化させることで電力を制御する。また、デルタ変調は元の信号の差分をパルスで表現する。デルタ変調の一種である  $\delta$  変調は、デジタルアンプなどでも使用されている。

本研究では、このうちパルス幅変調によって音声信号を伝送し (図 11)、PointSpeech を実装した。

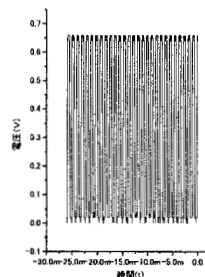


図 11 pointspeech による音声データ

#### 5. 考 察

##### 5.1 映像処理によるテレクリックの検出

本研究の大きな特徴の一つは、情報デバイスをテレクリックするにあたり、映像処理ではなく光信号を利用したことである。映像処理によって実世界を認識す

る研究は多数行われている<sup>5),6)</sup>。映像処理によってテレクリックを検出した場合、ジェスチャ入力的一种と考えられる。ジェスチャ入力は既に多数の研究がなされており<sup>7)~9)</sup>、特にPierceらの論文<sup>10)</sup>では、指差しを利用した同様のインタラクションが研究されている。しかし画像処理装置のみでテレクリックを検出した場合、対象物が何であるかを映像を利用して検知する必要が生じる。対象物が遠距離になる場合、対象物の映像の占める画素数は減少するため、対象物について得られる情報も減少する。また、画素数の少ないCCDなどを利用した場合も減少する。

これに対し、光信号を利用したテレクリックのシステムでは、画素数に依存して情報量が減少しないので、映像処理を利用した際の制限は生じない。従って、映像処理を利用するより遠距離での運用に適したシステム構成であると考えられる。

### 5.2 テレクリックの操作性

テレクリックは、3次元空間(実世界)の情報機器を選択する方法であるが、実際には映像上で情報機器をポイントしているの、操作に要する次元は2次元である。テレクリックを行う際のシステムの分解能は約3°であるが、ユーザが腕をいっばいに伸ばした際に人差し指の占める画角が約1°であると言われている。従って、ユーザが指差しを行う際に、人差し指1本半分の動きを分解できることから、ポインティングデバイスとして十分な分解能を持っていると考える事ができる。

この画角からピクセル数を計算すると、このCCDの画角は7.1°で、画素数は510×492であることから、単純計算で約25×25の分解能であることになる。Windows XPで使用されているアイコンは、48×48、32×32、24×24、16×16である。CCDの画素数はディスプレイの画素数の約半分であるので、ピクセルに換算した分解能は50×50ピクセルとなり、大きめのアイコン一つ分の分解能を持つと考えることが可能である。

また、実世界の情報デバイスを選択する方法として、レーザーポインタを利用した手法も研究されている<sup>11),12)</sup>。レーザーポインタを利用したインタラクションは、精度の高い選択が可能であるが、ポインタを保持する手首を3次元的に動かすことによって操作しなければならぬ。これに対して、テレクリックは2次元的な動作で操作することができる。従って、テレクリックはポインティングの際にユーザに要求される所作の精度が低いため、ユーザに対して負担を掛けないポインティング方法の一つであるとも考えられる。

### 5.3 発信器の構成による誤作動への耐性

本システムでは、光信号を遮断することでテレクリックを検知する仕組みを採用している。従って、ユーザと発信器の間を偶然何かが通過した場合、信号成分が生じ、誤作動が起きることは避けられない。本システ

ムでは、2つの発光素子から光信号を発生し、この内一つが遮断された場合にのみ信号成分が生じる仕組みである。従って、誤作動が生じるのは、2つの発光素子が発生された信号光のうち、1つのみが正確に遮断された場合である。こうした状況が生じるのは、ユーザから見て約1.5°程度の画角内に正確に遮断物が存在し、継続的にこの位置関係を保った場合のみである。偶然にこのような状況が生じる可能性は否定できないが、ユーザが意図的にこのような状況を作らない限り、こうした状況が継続する可能性は非常に低いと考えられる。

### 5.4 撮像装置の融通性

本研究の受信器は、原理的に撮像装置からの反射光を受光することができれば、撮像装置が何であっても動作する。例えば撮像装置が人間の眼球であった場合、ユーザが自らの視野上で「テレクリック」した対象物から光信号を受信する機能を持つ。これらの点は論文<sup>13),14)</sup>で取り上げられ、RealEyeCommunicatorとして紹介されている。

## 6. まとめ

本研究では、テレクリックと呼ばれるインタラクションを提案した。また、テレクリックされた送信器から音声情報を受け取る、PointSpeechと呼ばれる機材を作成した。PointSpeechは、カメラによって送信器を選択できる光受信器であり、50×50ピクセルほどの精度で実世界上の光送信器を選択し、情報を受信することができる。本研究では、光信号によって音声情報を送信し、カメラ上で送信器を「テレクリック」することで音声情報を受信することが可能な、PointSpeechと呼ばれるポインティングデバイス・光通信器を作成した。

## 参考文献

- 1) Mark Weiser. The computing for the 21st century. In *Scientific American*, pp. 94-104, 1991.
- 2) Roy Want, Kenneth P. Fishkin, Anuj Gujar, and Beverly L. Harrison. Bridging physical and virtual worlds with electronic tags. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 370-377, New York, NY, USA, 1999. ACM Press.
- 3) Roy Want. The magic of rfid. *Queue*, Vol.2, No.7, pp. 40-48, 2004.
- 4) Takuichi Nishimura, Yoshiyuki Nakamura, Hideo Itoh, and Hideyuki Nakashima. System design of event space information support utilizing cobits. *icdcsu*, Vol.03, pp. 384-387, 2004.
- 5) Jun Rekimoto and Yuji Ayatsuka. Cybercode: designing augmented reality environments with visual tags. In *Designing Augmented Reality*

- Environments*, pp. 1–10, 2000.
- 6) David Scott, Richard Sharp, Anil Madhavapeddy, and Eben Upton. Using visual tags to bypass bluetooth device discovery. *SIGMOBILE Mob. Comput. Commun. Rev.*, Vol.9, No.1, pp. 41–53, 2005.
  - 7) Andrew Wilson and Steven Shafer. Xwand: Ui for intelligent spaces. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 545–552, New York, NY, USA, 2003. ACM Press.
  - 8) Koji Tsukada and Michiaki Yasumura. Ubi-finger: Gesture input device for mobile use. In *Proceedings of APCHI 2002*, pp. 388–400, 2002.
  - 9) Ken Hinckley. Synchronous gestures for multiple persons and computers. In *UIST '03: Proceedings of the 16th annual ACM symposium on User interface software and technology*, pp. 149–158, New York, NY, USA, 2003. ACM Press.
  - 10) Jeffrey S. Pierce, Andrew S. Forsberg, Matthew J. Conway, Seung Hong, Robert C. Zeleznik, and Mark R. Mine. Image plane interaction techniques in 3d immersive environments. In *SI3D '97: Proceedings of the 1997 symposium on Interactive 3D graphics*, pp. 39–ff., New York, NY, USA, 1997. ACM Press.
  - 11) Jr. Dan R. Olsen and Travis Nielsen. Laser pointer interaction. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 17–22, New York, NY, USA, 2001. ACM Press.
  - 12) Brad A. Myers, Rishi Bhatnagar, Jeffrey Nichols, ChoonHong Peck, Dave Kong, Robert Miller, and A. Chris Long. Interacting at a distance: measuring the performance of laser pointers and other devices. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 33–40, New York, NY, USA, 2002. ACM Press.
  - 13) Yuichi Mitsudo and Ken Mogi. Real eye communicator: An eye-mediated real world pointing device. In *Proceedings of the First International Conference on Pervasive Computing (Short paper)*, pp. 112–118, Zurich, August 2002. In an informal companion volume of short papers.
  - 14) Yuichi Mitsudo. A model of axis-selectable optical data translation system using user's eye as a medium. In *Transactions of the Virtual Reality Society of Japan*, pp. 153–160, 2006.