

グループ生体情報による多視点映像コンテンツの作成

中村亮太* 市村 哲** 岡田謙一* 松下 温**

個人での映像制作・配信が一般的になってきた現在、より質の高い映像コンテンツの作成が求められている。本来、視聴者に分かり易く伝えるためには、放送型映像のように豊富な映像素材から作成される多視点での映像が望まれるが、多くの編集工程と時間を要するため、個人編集者に対する負担が大きいという問題がある。そこで著者らは、個人が容易に質の高い映像コンテンツを作成するために、編集時に行われる映像の選別を撮影時のカメラマンの心理状態から行うことで、編集作業の効率化を図ることを試みている。本稿では、撮影時におけるカメラマンの生体反応（脳波、筋電）から思考状態・カメラワークを推定し、シーンごとに重要度を判別することで複数アングルからの映像選択ならびにリプレイすべき重要なシーンの決定を行う手法について述べる。

Application of Reciprocal Vital Reactions

- Creating of the Video Contents -

Ryota Nakamura* Satoshi Ichimura** Kenichi Okada* Yutaka Matsushita**

Video cameras are becoming increasingly popular. We often edit videos with PCs. It is easy to produce video contents and distribute them on a network now. In order to create high quality video contents, we developed a system that used two or more cameras and tried creation of a multi-angle video. However, since much time is required for it, the burden on an individual is large. The purpose of our study is to change home videos automatically into more attractive images in an easy way. We aimed at the increase in efficiency of edit works, in order that an individual may create high quality video contents easily. In this paper, the state of mind and camera work are presumed from the vital reaction (EEG, EMG) of the cameraman at the time of video photogrammetry, and importance is distinguished on each scene on the basis of the result. The technique of selecting camera angles and decision of a replay scene in the videos are proposed.

1. はじめに

現在、デジタルビデオカメラが普及し、撮影した映像を PC で編集するための機器やソフトなどのマルチメディア環境が非常に充実したことで、プロだけのものではあった映像コンテンツの制作・配信が個人でも気軽に行えるようになってきた。また、企業や大学内においても遠隔テレビ会議収録システムや講義収録システムなど動画が扱われることが多くなっている。こういった状況の中で、より質の高い映像コンテンツの作成を行うため、CG を挿入することや編集ソフトを用いてテロップを表示させたり、映像の切り替え時にフェードなどの映像効果を適用したりすることが一般的になっているが、スポーツ中継映像のような多視点での映像コンテンツは少なく、1台のカメラのみで撮

影された単視点映像である場合が多い。1台のカメラ撮影では情報量が少なく、結果、制作者の意図が伝わりにくいことや映像を見る者にとって物足りないものになってしまうことがある。よって複数台のカメラを用いて、より多くの映像素材を手に入れることが望ましいが、そのような撮影システムは大規模でコストが高く、また撮影された多くの映像を編集する作業にも非常に手間がかかるため、個人での利用は困難であると考えられる。

著者らは以前、個人でも容易に多視点映像を扱うために多視点同期動画再生システム「Multi Video Player」(以下 MVP)の開発を行った[1]。MVP は、個別に撮影した映像から撮影日時情報を解析し、同時刻に撮影された映像を自動検出することで同期再生を可能としている。さらに個人撮影映像の質を向上させるための自動編集システムとして、スポーツ中継で多用されているリプレイの映像効果に着目したリプレイ自動生成システム「MindStudio」の開発を行った[2]。

* 慶應義塾大学大学院 理工学研究所

Graduate School of Science and Technology, Keio University

** 東京工科大学 コンピュータサイエンス学部

Tokyo University of Technology, School of Computer Science

MindStudio は撮影者の生体情報の変化から思考状態を予測し、興奮、集中状態になった映像シーンを自動的に検出することができる。そして、スポーツ中継映像の分析から得られた知識により適切なタイミングでリプレイを挿入・再生するシステムである。このシステムによって自動編集された映像は、ユーザーの注目したシーンがリプレイとして強調され、また、映像としてのリズムをつけることで面白みのある映像に仕上げることが可能となった。

本研究では、MVP の多視点機能と MindStudio の生体情報を利用した機能を拡張し、より魅力ある映像コンテンツに上げるためのシステム構築を行った。本稿では、撮影時におけるカメラマンの生体反応（脳波、筋電）から思考状態・カメラワークを推定し、シーンごとに重要度を判別することで複数アングルからの映像選択ならびにリプレイすべき重要なシーンの決定を行う手法を提案する。

2. 背景

2.1 個人映像制作における要求分析

デジタルビデオカメラユーザーへの調査[3]によると、動画を取り込んだことがあると答えたユーザーが映像を毎回編集する割合は 28%、ときどき編集するユーザーは 45%、特に編集をしないユーザーは 35%であった。編集をしない主な理由としてはビデオ編集に時間がかかる、面倒であることということ答えが大半であった。しかし、ビデオ編集をしたいと思っているユーザーは 59%と多く、撮影した映像を整理し、また面白みを加えたいと思っているユーザーが多いことが見受けられる。またネットワークを利用して特定の人物（家族、知人）に動画を公開したいと思っているユーザーは多い。以上のように映像コンテンツの制作・配信に対するニーズはあり、どのように映像編集が行える環境を作るかが課題であると考えられる。

2.2 映像制作における問題

前述したとおり、個人で撮影を行う場合、主に 1 台のカメラで撮影されることが多い。よって様々なアングルで被写体を捉えるような多面的な撮影が出来ないために障害が生じる。例えば、視聴者に見せたいものが死角に入ってしまうと伝えることが出来ない場合や、注目している被写体ばかり撮影してしまい、その場の雰囲気などの情報を伝えることができないことが考えられる。また、単視点映像は多視点映像のように映像の切り替えというリズムがなく、単調な映像になりがちである。放送型映像では、3 秒間以上同じアングルで放送することは視聴者を飽きさせてしまう恐れがあるため、映像の切り替えはたえず行われている。

次に編集時の問題であるが、編集作業は物理的には映像の並べ替えでしかないが、内容を整理し、ストーリーを見る人へ伝えるということは重要であり、また非常に複雑なものである。したがって映像素材が等しくても編集者の技量で作品の完成度が大きく左右される。個人では最適な映像構成を組み立てる技量がないことが多く、ただ闇雲に映像を繋ぎ合わせただけの編集になり兼ねない。また、市販されている編集ソフトなどにより容易に映像効果を導入することができるが、正しい適用方法を知らなければ逆効果になってしまう恐れがある。テレビ番組制作者の映像技法は、熟練した技術者による長年の経験によって培われた基本技法であり、映像文法と呼ばれる。様々な出来事を関連付け、ストーリーを作るために順序立てる編集を行っている。個人がプロの知識を獲得することも、また実際そのような編集を行うにしても、膨大な時間を要するものであり、かなりの負担になると考えられる。

2.3 従来の映像自動編集の問題点と関連研究

映像を制作するに当たって最も時間を要する作業は編集であり、作業の効率化が求められている。そのために従来、色情報、動き情報、形状情報、テキスト情報などの画像の特徴量や周波数、音声情報などの音の特徴量、またメタデータを付加したセグメンテーションやクラスタリングする手法が研究されている[4]～[8]。画像認識による映像インデキシングとして、料理番組を対象とした研究がある[4]。料理番組において、重要とされているシーンは調理をしている場面や素材の状態を示したときであり、その時の映像特性としては画面全体の動きが激しいという関連を見出した。これをオプティカルフローにより検出し、映像の要約を行っている。音声から重要場面を抽出する研究では、サッカーなどのスポーツ観戦において観戦者の歓声とスポーツの重要場面との間には深い関わりがあることを前提として、音声パワーの帯域別解析、重要場面に特有な音声パワーのパターン解析、重要場面と密接に関連する統計量について分析を行っている[5]。これらのようなパターン認識研究の多くは、対象とする映像に特化した処理を行うことで映像の意味内容を抽出するものであるが、精度良く要約映像を作成するためには、対象とする映像の種類を極めて限定する必要がある。個人撮影映像のように映像が構造化されていないものにこれらの手法を適用することは難しいと考えられる。

これに対して、放送映像のような明確な構造を持たないホームビデオをインデキシングする研究として

A.Girgensohn らは、編集支援ツール Hitchcock を提案し、半自動的に映像の処理を行い、ユーザーの負担を減らすことを可能としている[6]また、R.Lienhart は映像からタイムスタンプショットを階層に意味のあるまとまりにクラスタリングし、映像要約を行うアルゴリズムを提案している[7]。以上の研究は映像コンテンツを効果的に閲覧、編集、要約することが中心であり、魅力的な映像を作り上げるために映像効果を導入するといった映像の自動編集に関する研究は極めて少ない。本研究は如何に容易に魅力ある映像に自動編集できるかに焦点を当てている。

2.4 MVP の改善点と MindStudio の拡張

図 1 に示すように、我々が開発した MVP では、同時刻に撮影された映像の同期をとって再生させることが可能となっている。MVP により多視点からの映像を視聴することが可能となったが、実際に多視点映像をユーザーに視聴させたところ、複数の画面があるため、どこに注目すべきか迷ってしまうため煩わしさを感じるという意見があった。普段テレビで見慣れている映像は単画面で映像が切り替わるものであり、視聴者は放送局側によって与えられた映像を視聴しているため負荷を感じない。MVP においても放送型映像のように映像の自動切り替えが必要であると考えられた。しかし、映像を切り替えるためには当然それぞれのシーンにおいてどの映像を再生するか、編集時に複数の映像から選択する必要がある。映像の切り替えを自動的にある間隔でランダムに切り替えてしまえば、視聴者に見せたいものが表示されない場合や不快さを感じさせるような映像の乱れがあるものが含まれてしまう恐れがあるため適切ではないと思われる。しかしながら、映像の組み立て作業は非常に手間がかかり、また個人編集者の最も苦手とする作業であると考えられる。したがって必要な映像を的確に選択する自動編集システムが望まれる。そこで著者らは MindStudio で利用した撮影者の生体情報を映像の選択基準として行うことを試みた。また、MindStudio では単視点でのリプレイ映像であったが、複数アングルからの映像を連続的に再生するようにシステムの拡張を行った。

3. 脳波による重要シーンの自動検出

各映像シーンの重要度は人間の感性及び感情による部分が大きく、従来の画像、音声認識では抽出誤差が大きくなってしまおうと思われる。人間は、人や景色の変化などの視覚的刺激および心理的影響を受けることにより無意識的



図 1.MVP による多視点同期再生

に生体変化を起こすといわれている[8]。そこで著者らはビデオ撮影者の生体情報からシーンの重要度を判断できるのではないかと考え、脳波による特徴変化に着目し、重要シーンとの関連性を見出すことで自動的に抽出することを試みた。

3.1 脳波 (EEG) の特性

約 140 億個の神経細胞は、脳の活動状態や様々な刺激に関連してあるリズムを持った微弱な電位変化を示す。このような脳の電気活動を記録したものを EEG(脳波 ; Electro-encephalographic)という。一般に EEG は 1~50Hz の周波数で、約 10~100 μ V の振幅を示す。そして 14~30 以上[Hz]の波は速波と呼ばれ、意識レベルの高い状態や興奮した状態のときに現れることが報告されている[9]。また、この現象は、減衰とも呼ばれる[10]。年齢、意識レベル、外界からの刺激などによって影響されるため、はっきり脳波と精神状態の関係を決めることは困難であるが、おおよその心理状態を判定できる。

3.2 脳波測定装置と各脳波状態

図 2 に実験に用いた脳波測定装置(Mind Force)を示す。ヘッドバンド型をした脳波測定器は非常に小型かつ軽量であり、また無線でデータの送信を行えるため携帯性に優れている。着脱も数秒で行え、脳波測定者の自由度は大きい。各研究機関や大学で、リラクゼーションやコミュニケーションなどの研究が行われている[11]。また、ウェアラブルカメラにより長時間の日常生活を撮影した体験映像を要約する研究がある[12]。

図 3 に脳波測定実験によって得られた 4 つのパターンを示す。平常時においては気が散漫しているため約 7~20[Hz] 辺りまで一様に波形が見られる(A)。これが集中した状態になると全体的に脳波は減衰し、特に波の

減衰が顕著に現れる(b)。このとき 波は、他の周波数帯と比べてその減衰量は小さく、結果的にベータ波が目立つ。これが、前述した 減衰であると考えられる。(C)は人が興奮や緊張をした状態のときに突発的現れ、図のように 20[Hz]以上の 波領域の脳波が瞬時に増幅する。最後に(D)であるが、これは眼球運動や顔の筋肉の動きがあったときに現れる EMG (筋電: Electro-myographic) によるものである。以上この4つ基本パターンをもとに各シーンの精神状態を判断し、重要シーンの検出を行った。



図2. 生体情報測定器 (Mind Force)

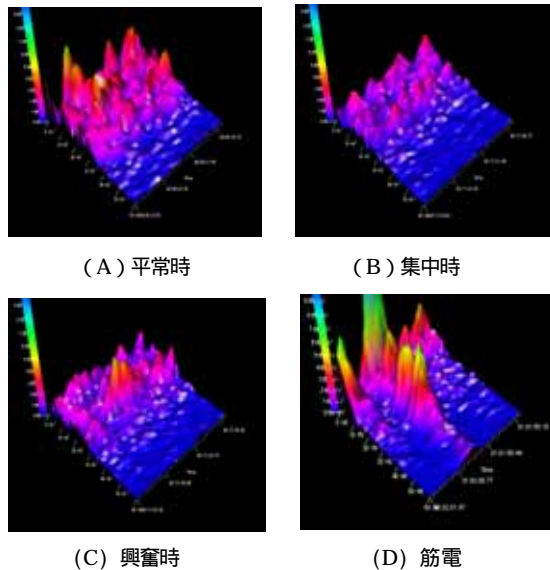


図3. 基本脳波パターンと筋電

3.3 脳波による注目シーンの検出実験



図4 脳波測定風景

脳波によって検出されたシーンと、実際に被験者が注目したシーンを比較し、脳波によるシーン抽出の有効性について検証する実験を行った。以下に実験方法ならびに結果を示す。被験者

(9名)に同一のスポーツ映像

(20 分間)を視聴させ、脳波測定を実施した。外界からの刺激で筋電起こり、脳波測定に影響を及ぼしてしまうと考えられるため、仕切りをして視界を遮り、ヘッドフォンの着用で、余分な音が耳に入らないよう考慮した。また、多少の顔の動作が予想されるので、カメラで撮影し、筋電による波形の変化でないかどうかの判断を行った。なお、視聴してもらった映像は比較的どのような人が見ても注目集めるような映像を用意した。そして、視聴後、被験者にどのシーンに注目したかを集計し、脳波によって無意識的に検出されたシーンと、実際に被験者が申告したシーンとの比較を行った。分析対象とした脳波の周波数は 波(7~12[Hz])、及び 波(13~30[Hz])であり、集中した状態: 波が 波と比較して相対的に増幅している(減衰)と興奮した状態: 波の突発的変位がおこった箇所を重要なシーンと見なした。次の表に結果を示す。

表1の結果により被験者全員について再現率が80%を上回る結果が得られた。これは被験者が注目したシーンの多くが脳波計によって無意識に起きた心の変化を検出することが出来たことを表している。このように高い再現率が得られた一方で適合率の低さが現れてしまった。表2の結果より被験者全員の平均適合率は56%とあまりよくない。被験者の注目シーンの申告漏れということも考えられるが、被験者が注目したシーン以外にも多くの箇所では脳波計が反応してしまい、このような結果になってしまった。再現率を高く見込むためには、ある程度多めにシーンの抽出をしなければならない。しかし、これは無駄に多くのシーンを抽出してしまう恐れがあり、適合率の悪化につながる。適合率を上げるよう検出閾値を上げつつ再現率を失わないよう調整しなければならないことがわかった。

表1. 脳波検証実験結果

	Bs	Sp	平均 再現率	平均 適合率
平均検出数	11.9	7.9	84%	56%
平均一致数	-	6.7		

$$\text{再現率} = \frac{\text{Bs, Sp 一致}}{\text{Bs 検出数}} \quad \text{適合率} = \frac{\text{Bs, Sp 一致}}{\text{Sp 検出数}}$$

Bs: 脳波の特徴量変化によって抽出されたシーン

Sp: 被験者が申告した注目シーン

3.4 撮影時の生体反応測定

被験者に約 10 分間のビデオ撮影をさせ、同時に MindForce によって脳波および筋電の変化を測定することで、撮影映像の動きと生体反応の比較を行った。図 5 に結果を示す。実線は撮影者の脳波反応より求められた興奮度を表し、グラフが上昇するほど、そのシーンに注目したことを表している。面グラフは顔の動きによる筋電反応であり、撮影中に激しく動いた場合グラフは上昇した。時間軸上の印は撮影時における急激な撮影方向の転換などにより映像の動きが激しく内容が理解できない瞬間を表している。

グラフの映像の動きと筋電反応を比較すると、カメラが激しく動くことにより映像が乱れた場合、75%と高い確率で筋電反応が現れることがわかった。また、筋電反応はそれほど上昇していないが、撮影者が、興奮状態になったとき、映像の乱れは確認されなかった。

以上の結果より、撮影者の生体反応を測定すれば、各シーンの重要度が推測でき、重要度の高い場合にはリプレイを行う、低い場合にはカメラアングルを切り替えるという判断を自動的に行うことが可能であると考えられる。今回この機能を新たにシステムに加えることとした。

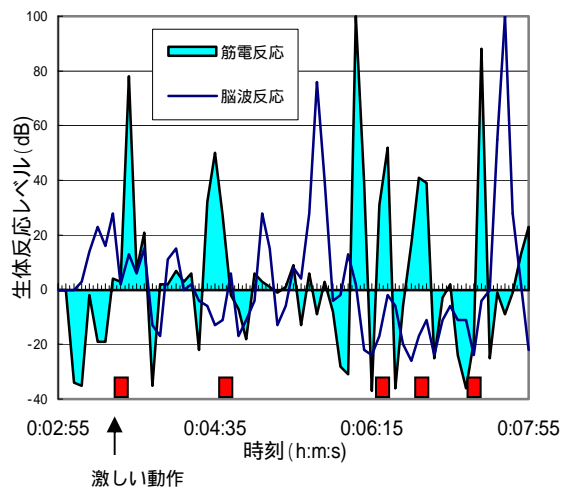


図 5．生体反応と激しい映像の動き

4．システムの構成

本システムでは、まず撮影映像と生体情報が Video-Analyzer, EEG-Analyzer それぞれにおいて解析される。Video-Analyzer では撮影映像のファイル名、撮影日時情報の登録が行われる。EEG-Analyzer では生体測定器によって測定されたデータが解析され、カメラマンの思考状態とカメラワークが推定される。それぞ

れの Analyzer によって得られたデータからリプレイ映像時刻、リプレイ挿入タイミングと映像切り替えタイミングが決定される。このときのリプレイ挿入タイミングは撮影一時停止時刻に相当する。最後に動画編集ツールにおいて動画ファイルの組み立て・映像効果の追加を行い映像コンテンツが完成する。

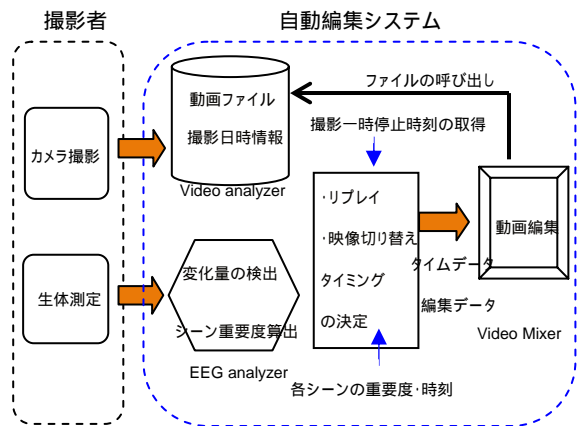


図 8．システム構成図

5．システム評価実験

5.1 実験方法

今回構築したシステムを用いて実際に多視点映像コンテンツの作成を行い、視聴者にアンケートを取り、システムの評価を行った。図に示すように体育館におけるバスケットボールの試合をハンディカメラと固定カメラ 2 台で撮影を行い、2 視点での映像を作成した。撮影者 A はハンディカメラで撮影対象に近い位置で撮影を行い、迫力のある映像や詳細が確認できる映像を撮影した。また、同時に MindForce により生体情報を取得した。撮影者 B は被写体の周り全体を上から固定カメラで撮影した。撮影者 A は小まめに撮影を一時停止するのに対し、B では撮影をする必要がない場合は除いて連続的に撮影を行うこととした。なお、2 台のカメラは時刻同期を取る必要があるため、事前にビデオカメラの時刻を合わせておく必要がある。図の脳波反応区間は撮影者が興奮状態にあった重要シーンの時間帯を表し、このシーンがリプレイとして撮影が一時中断された時間帯に挿入される。このリプレイでは、撮影者 A と B の撮影映像が連続的に再生され、注目したシーンがより強調されるような効果を持たせている。筋電反応区間は撮影者 A のカメラが激しく方向転換しながら撮影を行ったため乱れた映像になった区間である。この時間帯の映像は撮影者 B の映像に切り替えら

れるようになっている。

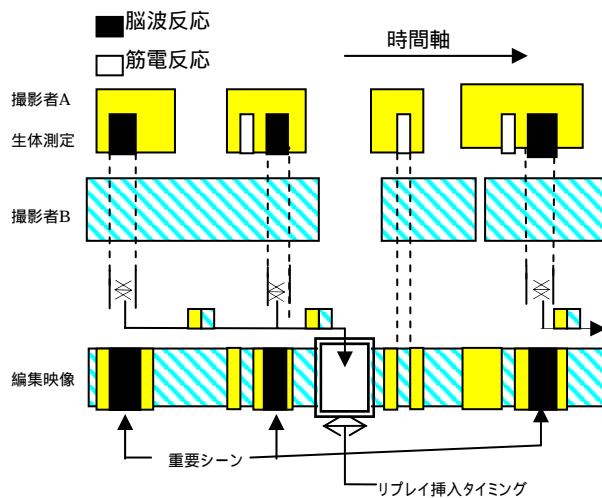


図9．多視点での映像構成

6.2 実験結果

表2に視聴した被験者10人の意見と感想をまとめた。画面の明るさや音声の大きさのずれが要因となって違和感がある視聴者が何名かいた。しかし、リプレイとカメラアングル切り替えの有効性を確認することはできた。

表2．視聴者アンケート（評価点1～5点）

被験者	評価点	被験者	評価点
A	3	F	4
B	2	G	4
C	3	H	3
D	4	I	4
E	3	J	3

平均評価 3.3

<意見>

- ・2つの映像の音声の大きさが異なっていたため、アングルが切り替わるたびに音声の強弱が発生してしまい、違和感があった。
- ・2つの映像で明るさが異なるため映像の繋がりがおかしい。
- ・若干不必要なシーンが含まれていた感じがした。
- ・リプレイはおもしろい。
- ・スポーツ中継映像に似た印象であった。

7. おわりに

本稿では個人撮影映像を容易に魅力度高い映像へ自動編集する一環として、多視点映像を目指し、撮影者の生体情報を利用したカメラアングルの切り替え、注

目シーンのリプレイを行った。実験により、2つの映像の組み立てを生体反応情報から容易に決定することができ、編集作業の負荷を大幅に減少することが可能となった。そして、多視点映像コンテンツとしての質も安定した映像とリプレイ映像効果という面白みを加えることである程度実現できたのではないと思われる。今後は評価実験結果を活かし、興味を引く映像づくりについて検討を重ねていくと同時に、複数人の生体情報から協調的な手法を研究する予定である。

参考文献

- [1] 市村, 宇田, 伊藤, 田胡, 松下 “遠隔実習教育のための多視点映像コンテンツ作成システム”, 情報処理学会 DICO 2003 論文集, pp.249-252 (2003).
- [2] 中村亮太, 市村 哲, 松下 温 “Mind Studio: 撮影者の生体情報を用いたリプレイ自動生成システムの試作” インタラクシオン 2004 論文集, pp.25-26(2004).
- [3] japan.internet.com 編集部, 2003年6月20日付記事 <http://japan.internet.com/research/20030620/1.html> 2003/09/01 にアクセス。
- [4] 三浦, 浜田, 井手, 坂井, 田中: “料理映像の特徴を利用した要約手法の検討” PRMU2002-22(2002-6), pp15-20
- [5] 佐野, 住吉, 柴田, 八木: “サッカーゲームにおける会場音からの重要場面抽出の検討” PRMU2001-166(2001-12) pp.87-92
- [6] A.Girgensohn, S.Bly, F.Shipma, J.Boreczky, and L.Wilcox, “Home Video Editing Made Easy? Balancing Automation and User”, In Human-Computer Interaction INTERACT, IOS Press, pp.464-471, 2001
- [7] Rainer Lienhart, “Dynamic Video Summarization of Home Video”, SPIE 3972 Storage and Retrieval for Media Databases 2000, pp.378-389, Jan.2000
- [8] 小杉幸夫, 武者利光 : 生体情報工学 電子情報通信工学 シリーズ 森北出版
- [9] 機械システム振興協会: “生体反応利用機器制御システムの開発に関するフィージビリティスタディ 報告書” 2003-03
- [10] 宮田洋 八木昭宏 他: “現代心理学シリーズ 2,6” 脳と心 “知覚と認知” 倍風館, 1996
- [11] 福井, 本田, 宮田, 重野, 岡田: “多地点遠隔コミュニケーションにおける思考状態ウェアネスの提供” CSCV 2003-4 -pp.33-38, 2003
- [12] 石島, 相澤: “ウェアブルによる長時間個人体験記録の編集 脳波を利用した映像の自動編集の試み” PRMU2000 - 158 (2001-01) pp85 92, 2001