

会議を撮影した動画メディアの思考状態インデキシングの提案

宮田 章裕[†] 福井 健太郎[†] 本田 研作[†]
重野 寛[†] 岡田 謙 一[†]

本稿では、会議を撮影した動画メディアの思考状態インデキシングを提案する。会議という知的作業において重要なのは「集中」や「関心」といった思考状態であるが、既存の手法はここに重点を置いていないものが多く、会議を長時間記録した動画メディアの中から「いつ」「誰が」「どのような思考状態にあったか」ということを判断することが難しい。そこで、本提案では、会議中の思考状態を表す指標として MS-Level (Mental State Level) を定義し、思考と深い関連があることが確認されている β 波を用いて MS-Level を求める。そして、導出した MS-Level を MS-Index (Mental State Index) という形式で表現し、会議を撮影した動画メディアをインデキシングする。我々は、提案概念を検証するために、動画メディア再生部と MS-Index から成るプロトタイプを実装した。プロトタイプは、会議を記録した動画メディアを思考状態という観点から参照できるように工夫されている。評価実験では、MS-Level の妥当性、MS-Index を利用して会議を撮影した動画メディアを参照することの有用性に関して良好な結果を得た。

A Proposal of Indexing Conference with Participants' Mental States

AKIHIRO MIYATA,[†] KENTARO FUKUI,[†] KENSAKU HONDA,[†]
HIROSHI SHIGENO[†] and KENICHI OKADA[†]

In this paper, we propose indexing conference with participants' mental states. Currently, one of the reasons why recorded movies of a conference do not represent participants' mental states is that existing ways do not place emphasis on recoding participants' concentration and interest. To address this issue, we define MS-Level (Mental State Level) derived from one's beta wave (a part of brain wave) as an indicator of depth of thinking, and index recorded movies of a conference with MS-Index (Mental State Index). We ran three experiments to evaluate our proposition using a prototype system consisting of a movie player and MS-Index, and conclude that our proposal will contribute to a better understanding of participants' mental states during conferences.

1. はじめに

我々は会議を行う際、その様子を後日参照するために、もしくは、会議に参加していない第三者に伝えるために、会議の流れや決定事項などを記録することが多い。従来はテキストとして記録することが多かったが、近年はマルチメディア技術の発達や記憶媒体の普及ともなあって動画メディアの取扱いが容易になったため、会議風景を撮影して動画メディアとして記録する機会が増えてきた。動画メディアによる記録ならば、表情、声色、ジェスチャなどテキストで表現することが難しい情報の記録も容易であり、会議の様子や雰囲気などを把握する際には非常に役立つ。

しかし、会議という知的作業において重要なのは「集中」や「関心」といった思考状態であるが、会議を長時間記録した動画メディアの中から「いつ」「誰が」「どのような思考状態にあったか」ということを判断することは容易ではない。

そこで、本研究では、脳波の成分である β 波を利用して思考状態を記録し、会議を撮影した動画メディアの索引とすることを提案する。会議中の思考状態を表す指標として MS-Level (Mental State Level) を定義し、思考と深い関連があることが確認されている β 波を用いて MS-Level を求める。そして、導出した MS-Level を MS-Index (Mental State Index) という形式で表現し、会議を撮影した動画メディアをインデキシングする。

MS-Index を動画メディアの索引として利用することにより、会議中の各シーンにおける各参加者の思考

[†] 慶應義塾大学大学院理工学研究科
Graduate School of Science and Technology, Keio University

状態が分かりやすくなる。また、参加者が特定の思考状態（集中、関心など）にあったシーンを見つけやすくなるなど、アクセシビリティが高い動画メディアになる。

以降 2 章で、会議中に発生する情報の記録について述べる。3 章で会議を撮影した動画メディアの思考状態インデキシングを提案し、4 章でプロトタイプの実装について述べる。5, 6, 7 章では提案概念を評価するための実験について述べ、最後に 8 章で本研究の結論と今後の展望について述べる。

2. 会議中に発生する情報の記録

会議とは、人間が行う協調作業の代表例であり、情報や意見の交換、問題の提起や解決、相対する意見の調整などが行われる。そして、その様子を後日参照するために、もしくは、会議に参加していない第三者に伝えるために、会議の流れや決定事項などを記録することが多い。従来はテキストとして記録することが多かったが、近年はマルチメディア技術の発達や記憶媒体の普及とともに動画メディアの取扱いが容易になったため、会議を撮影して動画メディアとして記録する機会が増えてきた^{1),2)}。

テキストが発言内容や決議事項などのパーバル情報を正確に記録することに秀でているのに対し、動画メディアはテキストで記録することが難しいノンバーバル情報を記録することに秀でているため、会議の様子や雰囲気や直感的に把握する際には非常に有用である。

しかし、会議を長時間撮影した動画メディアからは、重要なシーンなどを素早く探すことが難しい。ここで、会議中に発生したイベントや参加者の行動などの情報を抽出し、その情報を元に動画メディアをインデキシングすれば、効率良く動画メディアを参照できると考えられる。

以降 2.1, 2.2 節では会議・コミュニケーション中にあふれる様々な情報を記録・利用している研究について述べる。

2.1 ノンバーバル情報の利用

人が会議を行う際、そこには表情、ジェスチャ、声色、雰囲気などのノンバーバル情報が多く含まれている。たとえば、人が誰かを説得したいときは相手をじっと見つめて話し、感情が高ぶればジェスチャが大きくなり、声も大きくなる。重要な話題になれば参加者は真剣な顔つきをするし、会議が一段落すれば場は和やかな雰囲気に包まれる。このように、会議においてノンバーバル情報は重要な役割を果たしており、それらを記録・利用する試みも数多く行われている^{3)~6)}。

Meeting Browser³⁾ は、各参加者が発言している時間帯にインデックスを付加した発話ログと、簡単なテキスト議事録を作成する機能を持っている。Hindusらの研究⁴⁾でも同様の発話ログを作成している。また、Picard は声の抑揚やピッチから発話者の恐れ、悲しみ、幸福などを測定する試みを行っている⁷⁾。表情については、下田らが画像処理とファジィ推論を利用した表情認識方法を提案している⁵⁾。仮想オフィスシステム Valentine⁶⁾ では、キーボードやマウスの使用頻度と椅子を動かす頻度から参加者の集中度を測定している。

2.2 脳波情報の利用

脳波とは、脳の活動にともなって頭皮上に生じる電位のことであり、人間からつねに発生し続けている生体情報の 1 つである。脳波には、眼球運動などにより発生するノイズ (0~1 Hz)、 δ 波 (1~4 Hz)、 θ 波 (4~8 Hz)、 α 波 (8~12 Hz)、 β 波 (12~40 Hz) などの成分が含まれており、なかでも β 波は思考を要する作業を行うときに強く出現し、思考を要しない作業時にはあまり出現しないという特徴がある^{8),9)}。

これを利用して、脳波情報から導出した参加者の思考の状態を互いにアウェアすることがコミュニケーションに与える影響を調査した研究がある^{10)~13)}。また、ウェアラブル機器で記録した個人体験映像をインデキシングする際に映像と同時記録した脳波情報を利用する研究や、スポーツ中継のダイジェストを生成する際に視聴者の脳波情報をキーの 1 つとして利用する研究もある^{14),15)}。その他の研究例としては、感性スペクトル解析法¹⁶⁾ があげられる。この手法では、脳波情報の中から各感情に対応する特定パターンを抽出して感情解析を行っている。

3. 会議を撮影した動画メディアの思考状態インデキシングの提案

知的作業である会議において特に重要なのは参加者の思考状態、つまり、会議への集中の度合いや議題への関心の具合などである。しかし、会議を長時間撮影した動画メディアを参照する際、集中や関心の変化が明示的に記録されていないので、各シーンにおいて参加者がどのような思考状態にあるか分かりにくい。そこで、我々は会議を撮影した動画メディアの思考状態インデキシングを提案する。

3.1 β 波情報の利用

思考状態を解析する際に β 波情報を利用する。その理由は以下のとおりである。

(1) 思考状態と密接に関係している。思考を要する

作業時に強く出現し、思考を要しない作業時にあまり出現しない^(8),9)。

- (2) 表情やジェスチャなどには表れないような内面的な状態を推測することができる。
- (3) 主に額部で検出できるので、電極数が少ない簡易脳波計を額部に装着するだけで測定が可能である。
- (4) ジェスチャや音声とは違い、つねに発生し続けている情報である。

関連研究として、 β 波が強く出現する状態（つまり、 α 波が減衰する状態）を検出して、小型カメラを着用して動き回りながら記録した日常生活の映像をインデキシングする際のキーとするものがある⁽¹⁴⁾。それに対して、本稿では、脳波計測の対象となる活動はあくまで会議である。会議には様々なスタイルがあるが、総じて参加者が動きまわることは少なく、体の動きによって脳波にノイズが乗る可能性が少ない。発声や表情変化などでノイズが生じるが、計測対象を「動きまわらない知的作業」に限定しているため、ノイズのパターンも限られており（多くの場合、低周波数帯に瞬間的に生じる）、フィルタリングして除去することも難しくない。つまり、本研究では、脳波へのノイズの影響が比較的少なく、パターンが限定されているため取り除きやすいといえる。

また、スポーツ中継の視聴者の α 波が減衰する状態を検出して、視聴者が興味を覚えるようなシーンの抽出を試みる研究がある⁽¹⁵⁾。それに対して、本稿で脳波計測の対象となっている会議は、自ら参加し、考え、発言し、議論を交わすという積極的な思考が求められる活動である。連続的に動画を見続けるという行為を行うと、前頭葉の活動が鈍ってしまう現象が報告されているが⁽¹⁷⁾、他者とさかんにコミュニケーションを行う会議という作業中ならば前頭葉は活発になる。 β 波は主に前頭葉で発生しているので、会議という場面を前提としている本研究では β 波を検出しやすく、「どの程度頭を働かせているか」という思考状態も検出しやすいといえる。

3.2 MS-Level (Mental State Level)

思考状態を数値化するために、最新 n サンプルの β 波データの中で、その強度が閾値を超えたものの割合 (0~1) を求め (図 1)、これをその瞬間の MS-Level (Mental State Level) と定義した。閾値は各参加者が事前にキャリブレーションを行い設定しておく。キャリブレーションは以下のように、リラックスした状態を基準にして行う。

- (1) 閾値を高めに設定して MS-Level の測定を開始

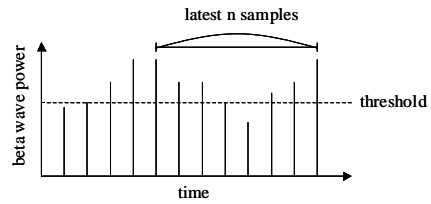


図 1 MS-Level の概念

Fig. 1 The concept of MS-Level.

する。

- (2) 静かな部屋で目を閉じて 3 分以上リラックスする。
- (3) MS-Level が 0.1 以下で 1 分以上安定する範囲で、閾値を徐々に下げることができるだけ低めに設定する。

MS-Level を導出する際に過去の n サンプルを利用した理由は、多くの場合思考は一瞬間だけで行われるものではないと考えられ、また、思考以外に β 波に影響を及ぼすノイズ（眼球運動などによるもの）は瞬間的なものが多いからである。このように、思考を幅のあるプロセスととらえて、各瞬間の思考状態の導出に過去の脳波データを利用している点も、前述の他研究^(14),15) との差異である。

n の値が大きいくほどノイズの影響が少ないが、短時間における思考の変化をとらえにくくなる。なお、「驚き」や「閃き」などのように思考にも瞬間的な要素があると思われるが、今回はこれらの要素は考慮せず、ある程度長期的な思考状態の変化を取り扱うことにした。

3.3 MS-Level の相対化

3.2 節で導出した MS-Level は各参加者の思考状態を表しているが、それを複数の参加者間で比較することは不適切である。なぜならば、 β 波強度の絶対的な数値には個人差があり、そこから導出した MS-Level にも個人差が含まれてしまうからである。

そこで、複数の参加者間で MS-Level の比較が可能になるように MS-Level の相対化を行った。手順は以下のとおりである。

- (1) 参加者ごとに MS-Level の平均値と標準偏差を求め。
- (2) 平均から標準偏差の 3 倍以上離れているデータ（一般に外れ値であるといわれている）を参加者ごとに取り除く。
- (3) 参加者ごとに最小値 0、最大値 1 になるようにデータを標準化する。

なお、相対化の作業は、会議がある程度長時間であ

ることを前提としている．なぜならば、思考の様子が平坦な場合、つまり、すべての参加者にそれぞれ「ほとんど思考を行っていない時間帯」と「かなり思考している時間帯」の両方が存在していない場合に相対化を行うと不都合が生じてしまうからである．たとえば、参加者 A はつねに何も考えておらず、参加者 B はつねに深い考えごとをしていたような短時間の会議で相対化を行ってしまうと、両者の MS-Level はどちらも同じように 0 から 1 の範囲に標準化されてしまい、2 人の思考の違いを表せなくなってしまうおそれがある．

それに対して、長時間の活動であれば、各参加者にはそれぞれまったく思考を行わない時間帯と、深く思考を行っている時間帯の両方がある可能性が高く、上記のような不都合は生じない．

3.4 場の雰囲気

各参加者の MS-Level を参照することで個々の思考状態が把握できるが、場合によっては以下のような問題が生じてしまう．

- (1) 参加者達の総体的な思考状態が分かりにくい．
- (2) 個人のプライバシーが侵される可能性がある．

参加者の人数が多いときに (1) の問題は深刻である．たとえば、大規模な講演において聴衆の反応を分析したい場合には、個々の人物の反応よりも聴衆全体としての反応の方が重要だからである．また、MS-Level が思考の内容まで表しているわけではないとはいえ、思考状態の変化を他人に知られることは抵抗がある場合もあるので (2) のような問題がないとはいえない．

そこで、これらの問題を解決するために複数の参加者（もしくは、全参加者）の MS-Level を平均して表現する手法を考案し、この指標を「場の雰囲気」と定義した．

3.5 MS-Index

MS-Level を利用して、会議を撮影した動画メディアの時間軸に沿った索引を作成し、この索引を MS-Index と定義する．

このとき、MS-Level の数値をそのまま文字表示して索引にする方法は不適切である．文字表示は比較的省スペースで正確な数値を伝えることができるが、数値が時間変化する様子や全体の分布状況を読み取るには時間がかかってしまうからである．逆に、グラフ表示であれば直感的に変化の様子や分布状況を把握することができるが、表示に大きなスペースを占有してしまうという欠点がある．

そこで、本研究では MS-Level の大きさに応じて m 段階に色の濃淡を付けたラベルを時間軸に沿って並べて MS-Index を作成する手法をとった．このとき、

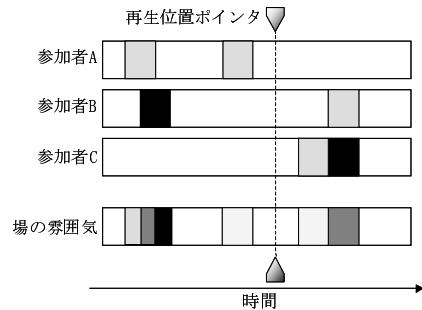


図 2 MS-Index の概念図

Fig. 2 The concept of MS-Index.

MS-Level の数値が大きいほど濃い色で表示した．文字表示とは違い、色の濃淡で表示された情報は直感的で素早く理解できる．また、グラフ表示とは違い、小さいラベルを並べるだけなので表示に大きな面積を必要としない．

図 2 に示すのが MS-Index の概念図である．再生位置ポインタは動画メディアの再生位置を示しており、再生が進むに左から右へと移動していく．再生位置ポインタが指し示す位置にあるインデックスを見ることで現在再生中のシーンにおける各参加者の思考状態、および、場の雰囲気が把握できるようになっており、逆に、インデックスを目安に再生位置ポインタを動かして任意のシーンにアクセスすることも可能である．

4. プロトタイプの実装

4.1 MS-Level の導出

参加者の脳波を計測する際には IBVA Technologies 社の簡易脳波計 IBVA を利用した．IBVA は、送信機、受信機、および、ヘッドバンド（3 電極付き）から構成されており、参加者はヘッドバンドを頭部に巻き付けるだけで脳波計測が可能である．

本研究では脳波の中から β 波成分を抽出するために、脳波計から得られた生データを高速フーリエ変換して 12 - 40 Hz 成分（2.2 節参照）のデータのみを取得した．今回の実装では高速フーリエ変換された結果データを約 0.25 秒ごとに得ることができ、周波数分解能は約 0.94 Hz となった．そして、約 0.25 秒ごとの結果データに含まれる各周波数帯（12 - 40 Hz 帯を 0.94 Hz ごとに分解した約 30 帯域）の強度の平均値を求め、これを 1 サンプルの β 波データとした．

4.2 プロトタイプの作成

3.5 節で説明した MS-Index を利用して、思考状態インデキシングを施した動画メディアを参照するためのプロトタイプを作成した．図 3 に示すように、プ

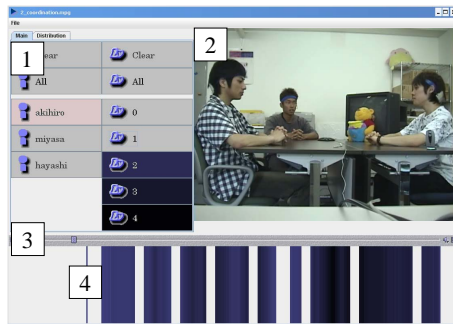


図 3 プロトタイプ

Fig. 3 A prototype system.

プロトタイプは、(1) 表示対象選択ボタン、(2) 動画メディア再生パネル、(3) メディアコントローラ、および、(4) MS-Index 表示部から成る。

(1) 表示対象選択ボタン

対象者選択ボタンと MS-Level 選択ボタンから成る。対象者と MS-Level を選択することで対応する MS-Index が表示される。複数の対象者を選択することで「場の雰囲気」を表示することもできる。

(2) 動画メディア再生パネル

会議の様子を記録した動画メディアを表示する部分である。

(3) メディアコントローラ

メディアの再生、停止などを行う部分である。シークバーはメディアの再生位置を示している。シークバーを動かして任意の再生位置にアクセスすることもできる。

(4) MS-Index 表示部

MS-Index を表示する部分であり、メディアコントローラの下に配置されている。シークバーの真下にある MS-Index を参照することで、そのとき再生しているシーンの MS-Level を表示することができる。

5. 実験 1：MS-Level の導出実験

5.1 実験内容

被験者 20 人 (21 ~ 25 歳の学生, 男 17 人, 女 3 人) に対して思考を要するタスク (英語リスニングテスト) と思考を要しないタスク (何も考えずにリラックス) を課した。タスクの時間はどちらも約 4 分間であり、タスク中の MS-Level の平均値を比較した。このとき、思考を要するタスクに英語リスニングテストを選んだのは、思考に最も関連が高いとされている β 波が、様々な学カテストの中でも特に言語テストを行う

表 1 実験 1 の実験結果

Table 1 The result of the experiment 1.

| | 思考を要する | 思考を要しない |
|----------------|--------|---------|
| < $n = 1$ > | | |
| MS-Level の平均値 | 0.649 | 0.301 |
| MS-Level の標準偏差 | 0.201 | 0.159 |
| < $n = 25$ > | | |
| MS-Level の平均値 | 0.648 | 0.256 |
| MS-Level の標準偏差 | 0.187 | 0.151 |
| < $n = 50$ > | | |
| MS-Level の平均値 | 0.638 | 0.246 |
| MS-Level の標準偏差 | 0.179 | 0.148 |
| < $n = 75$ > | | |
| MS-Level の平均値 | 0.610 | 0.270 |
| MS-Level の標準偏差 | 0.171 | 0.121 |
| < $n = 100$ > | | |
| MS-Level の平均値 | 0.618 | 0.270 |
| MS-Level の標準偏差 | 0.155 | 0.117 |
| < $n = 125$ > | | |
| MS-Level の平均値 | 0.615 | 0.278 |
| MS-Level の標準偏差 | 0.151 | 0.109 |

(結果はすべて被験者 20 人の平均値, n は利用する最新サンプル数)

ときに多く発生するという知見があるからである⁸⁾。

実験は以下のような条件で行った。

- (1) 脳波計に対する不快感や実験に対する緊張などを和らげるため、被験者が脳波計を装着してから実験を開始するまで十分に時間をおいた。
- (2) 被験者が「思考」について必要以上に意識することを避けるため、被験者には何を計測するかいっさい知らせず、スピーカから聞こえてくる台詞を真似るタスクをダミー実験として織り交ぜた。
- (3) 被験者を 2 グループに分けてタスクを行う順番をグループ間で変えた。

また、実験は、MS-Level を導出する際に利用する最新サンプル数 (3.2 節参照) が $n = 1, 25, 50, 75, 100, 125$ の各場合について行った。それぞれ約 0.24, 6, 12, 18, 24, 30 秒分にあたる。

5.2 実験結果と考察

各タスクにおける全被験者の MS-Level の平均値は表 1 のようになった。なお、3.3 節でも述べたように、異なる被験者同士の MS-Level の平均をとる際には、事前に各被験者の MS-Level を相対化してある。その際、同節で触れたように、「思考を要するタスク」や「思考を要しないタスク」のどちらかだけを行っている時間帯で相対化してしまうと問題が生じてしまうので、一連のタスクを行っているすべての時間帯のデータを利用して相対化を行った。

表 1 を見ると、おおむね n の値が大きくなるにつれて思考を要するタスク中の MS-Level は徐々に小さ

くなり、思考を要しないタスク中の MS-Level は徐々に大きくなる傾向があることが分かる。また、各タスク中の MS-Level の標準偏差は n の増加にともない徐々に小さくなっている。

これは、ある瞬間の MS-Level を導出する際に利用する過去データの時間を長くするほど短時間の思考の変化の重みが減って、MS-Level が思考をしても大きくなりやすく、思考をしなくても小さくなりやすいためだと思われる。つまり、利用する過去サンプルが多いほど MS-Level が敏感に思考の変化を反映できなくなると判断できる。

このことから、利用する過去サンプル数はある程度少ない方が MS-Level が的確に思考状態の変化を反映できると思われる。しかし、表 1 をよく見ると、思考を要しないタスクの場合は n が 1 から 50 までは MS-Level の平均値が下がり続け、それ以降は上昇し続けていることが分かる。これは、 n の値があまりにも小さいと眼球運動などの思考以外に β 波を高めてしまう瞬間的な現象が悪影響を及ぼし、MS-Level を高めてしまっているためであると思われる。

よって、ノイズの影響が少なく、かつ、思考の変化を見失わない長さとして、過去 50 サンプル（約 12.5 秒分のデータ）を利用して MS-Level を導出することにし、6 章の実験 2 でもこの条件を用いた。

$n = 50$ のとき、思考を要するタスク中と思考を要しないタスク中の各被験者の MS-Level の平均値の集合の間で Wilcoxon 符号順位検定を行うと棄却率は $p = 0.0000820$ ($N = 20, p < 0.01$) となり、この 2 群の間には有意水準 1% で有意に差があることが分かる。よって、タスクに要する思考の度合いに応じた MS-Level が導出できていたと判断することができる。

また、思考を要するタスクの方が思考を要しないタスクに比べて標準偏差の平均値が大きい傾向が見られたが、これは、被験者によって英語の能力が異なっているために思考の度合いにばらつきが生じたことが一因であると思われる。

6. 実験 2：会議を撮影した動画メディアの分析実験

6.1 実験内容

被験者 24 人（21～25 歳の学生、男 20 人、女 4 人）に対して 2 種類の動画メディアを分析するタスクを課した。どちらの動画メディアも図 3 のようなインタフェースを利用して参照することができるが、表示されるインデックスが片一方は MS-Index、もう一方は発話 Index となっている。

MS-Index は、5 章の実験 1 で導出したように $n = 50$ サンプルの過去データを利用して MS-Level を求め、 $m = 100$ 段階の濃淡を付けた色ラベルで表示した。また、発話 Index とは 2.1 節で紹介したように、各参加者が発話していた時間帯を抽出・インデックス化した手法である^{3),4)}。発話 Index の作成に関しては、ビデオ撮影時に各参加者に装着したマイクに音声入力がある時間帯を抽出し、MS-Index と同様に時間軸に沿って発話していた時間帯を各参加者ごとにラベルで表示した。つまり、発話 Index とは音声入力の有無を 2 色のラベルで表示したものである。

動画メディアには、1 つのグループが会話する様子が撮影されている。その際、2 種類のインデックスで解析を行うために同じ形式で内容の異なる会話を 2 パターン用意した。会話は以下に示す順番で進んでいく。

- (1) 3 人の人物が円卓を囲んで座る。
 - 円卓の上にはカードが 7 枚置いてある。
 - 各カードにはテーマが 1 つずつ書いてある。
 - テーマは音楽、スポーツ、宗教など多岐にわたる。
 - 各人物が関心を持っているテーマが 2 つずつ含まれている。
 - 1 つのテーマに対して複数の人物が関心を持つことはない。
 - 1 つだけ誰も関心がないテーマが含まれている。
 - 事前に注意深くテーマを選んだため、各人物のテーマに対する関心の有無は本物である。
 - (2) 1 人がカードをめくり書いてあるテーマについて語り始める。
 - カードは伏せてある。
 - カードは無作為に選ぶ。
 - (3) そのテーマについて 3 人が語り合う。
 - 約 1 分間語り合う。
 - 会話に台本はなく自由に語り合う。
 - 明示的に関心の有無を示す発言はしない。
 - (4) (2)～(3) の手順をカードがなくなるまで繰り返す。
 - カードをめくる役割は順番に交代していく。
 - 対談は合計約 7 分間で終了する。
- 被験者には、上記のような対談を分析するタスクを課した。詳細は以下に示すとおりである。
- (1) 被験者の目標
 - 各人物が関心を持っていたテーマを当てる。
 - 誰も関心がなかったテーマを当てる。

表 2 実験 2 の実験結果
Table 2 The result of the experiment 2.

| | 発話 Index | MS-Index |
|-----------|----------|----------|
| 正解率の平均値 | 0.518 | 0.685 |
| 解答不能率の平均値 | 0.065 | 0.006 |

(結果は被験者 24 人の平均値)

(2) 利用できる手段

- 動画メディア (音声付き)
- 各テーマの内容と話される順番の一覧表
- インデックス (MS-Index/発話 Index)
- メディアコントローラを利用した動画メディアの各再生位置への自由なアクセス

(3) 制約

- 分析の制限時間は 4 分間 (動画メディアの長さは約 7 分)。
- 「テーマを語り始めたから関心がある」という一般的な認識が利用できない (テーマは伏せられたカードをめくった人から語り始められるから)。
- 「好き、嫌い」などの明示的に関心の有無を示す発言がビデオ中の会話で使われていない。

また、実験の公平性を期すため、被験者を 6 人ずつ 4 グループに分け、2 パターンの動画メディア、および、それらを分析するための 2 種類のインデックスの組合せを各グループ間で変えて実験を行った。

被験者は 7 つのテーマそれぞれに対して関心がある人物、もしくは、誰も関心がないことを答え、その正解率を集計した。

6.2 実験結果と考察

表 2 を見ると、MS-Index を利用した方が、発話 Index を利用した場合よりも高い正解率を出していることが分かる。つまり、MS-Index を利用した方が分析の正確性が高いという結果になった。この差が有意なものであることを確認するために、MS-Index を利用した場合の各被験者の正解率の集合と、発話 Index を利用した場合のそれとの間で Wilcoxon 符号順位検定を行ったところ、棄却率は $p = 0.0191$ ($N = 24$, $p < 0.05$) となり、この 2 群の間には有意水準 5% で有意に差があることが分かる。よって、この実験のようにコミュニケーション中の関心や集中について分析する場合、発話 Index よりも MS-Index の方が適していると判断できる。

さらに、表 2 からは、MS-Index を利用した場合の解答不能率 (「答が分からない」と解答した確率) が発話 Index を利用した場合の 1/10 以下であることも

表 3 実験 2 のアンケート結果
Table 3 The result of the questionnaire (experiment 2).

| | 評価の平均値 (低: 1 - 5: 高) | |
|----------------|----------------------|----------|
| | 発話 Index | MS-Index |
| 関心の様子が分かった | 2.42 | 4.21 |
| 意見が飛び交う様子が分かった | 3.96 | 2.46 |
| インデックスを分析に利用した | 3.17 | 4.46 |
| インデックス化が有効と感じた | 2.75 | 4.42 |

(結果は被験者 24 人の平均値)

分かる。被験者に解答不能になった理由を尋ねると、「時間が足りなくて判断が間に合わなかった」という意見を多く得た。解答不能率が大幅に少ないということから、MS-Index を利用した方が効率良く短時間で議事録の分析を行えるといえる。

表 3 に示すのは、議事録の分析を行ってもらった際に被験者に対して行ったアンケートの結果である。

- 関心の様子が分かった
MS-Index を利用した方が評価が高かった。これは、MS-Index の方がよりの確に思考状態を示していたためであると考えられる。逆に、発話 Index からは発話の有無や発言頻度などが分かるが、それらと思考状態は必ずしも一致しないのでそこから関心などを判断することは難しかったと思われる。
- 意見が飛び交う様子が分かった
発話 Index を利用した方が評価が高かった。これは、発話 Index が発話のあった時間帯を示しているため、各参加者の発話頻度などを判断する有力な手段となりえたためであると思われる。
- インデックスを分析に利用した
MS-Index を利用した方が評価が高かった。これは、今回のタスクが「各テーマに関心がある人物を当てる」というものであったため、思考状態が分かりやすい MS-Index を分析に利用した被験者が多かったのだと思われる。
- インデックス化が有効と感じた
MS-Index を利用した方が評価が高かった。これは、人物の関心などの思考状態を把握する際には、MS-Index を利用した方が有効だと認められた結果であると思われる。

また「MS-Index と発話 Index を組み合わせて分析を行いたい」という意見を述べる被験者もいた。これは的確な指摘であり、思考状態と発言頻度の両方を把握することでより多角的で正確な分析ができると考えられる。今後は両者のインデックスを組み合わせた分析手法を検討していく必要がある。

7. 実験 3：会議シーンの重要度判定実験

7.1 実験内容

被験者 24 人 (21 - 48 歳の研究員・学生, 男 22 人, 女 2 人) に対して, 会議を撮影した動画メディアの各シーンについて, 3 通りの方法で重要度を判定するタスクを課した. 詳細は以下のとおりである.

7.1.1 会議を撮影した動画メディアの準備

会議には, 創造会議, 決定会議, 調整会議, 伝達会議などのタイプがある⁽¹⁸⁾, 実験用の題材としては「創造会議」と「決定会議」「調整会議」の 3 パターンを撮影した. なぜならば, 本稿では会議中の思考状態に主眼を置いているので, 創造的問題解決のためにきわめて知的な発言が要求される「創造会議」と, 組織・集団の意思決定や意見調整に関して十分な吟味や合理的な判断が要求される「決定会議」「調整会議」がふさわしいと考えたためである. 具体的な議題は以下のとおりである. なお, より現実的な場面での有用性を評価するため, 会議を行う際には台本や特別ルールなどをいっさい設けず, 自然なスタイルで行った. 会議参加者は 3 人である.

- (1) 創造会議
研究室で行うイベントを考案する.
- (2) 決定会議
喫煙スペースを設置すべきかどうか決定する.
- (3) 調整会議
会議参加者全員のスケジュール調整をする.

7.1.2 会議参加者による重要度判定

会議を撮影した動画メディアの長さはすべて約 5 分になり, この動画メディアを会議の流れに沿って 10 シーンに区切った. そして, 会議終了直後に会議参加者全員で協議を行い, あらかじめ各シーンの重要度を「重要」「やや重要」「普通」「やや不要」「不要」の 5 段階で判定しておいた.

7.1.3 被験者による重要度判定

被験者 (会議に参加していない第三者) には以下の 3 通りの方法で, 7.1.2 項で会議参加者が区切った各シーンの重要度を判定してもらった.

- (1) 方法 1
動画メディアだけを見て判定.
- (2) 方法 2
動画メディアと発話 Index を見て判定.
- (3) 方法 3
動画メディアと MS-Index を見て判定.

なお, MS-Index, 発話 Index の作成方法は実験 2 と同じである. 各シーンの重要度を判定する際は, た

表 4 実験 3 の実験結果

Table 4 The result of the experiment 3.

| | マッチング率 |
|-------------------------|--------|
| 方法 1 (動画メディアのみ) | 0.488 |
| 方法 2 (動画メディアと発話 Index) | 0.552 |
| 方法 3 (動画メディアと MS-Index) | 0.644 |

(「方法 1~3」の結果は被験者 24 人の平均値)

たとえば, 「方法 1 では創造会議, 方法 2 では決定会議, 方法 3 では調整会議」といったように各方法で対象となる動画メディアを変えるようにした. さらに, 公平性を期すため, 被験者ごとに各方法で視聴する動画メディアの組合せを変えた. 制限時間に関しては, 各方法で 10 分ずつとした.

7.2 実験結果と考察

実験結果を定量的に評価するため, 会議参加者が判定した重要度 (7.1.2 項参照) と被験者が各方式で判定した重要度 (7.1.3 項参照) の「マッチング率」を集計した. 「マッチング率」とは, 「会議参加者による判定」と同じ重要度に判定されていたシーンの数を全シーン数 (10 シーン) で割ったものである. 会議の終了直後に当事者である会議参加者が全員で協議して定義した重要度は, 会議の様子をかなりの確に表していると考えられるので, 「会議参加者による判定」との「マッチング率」が高いほど, 会議の様子を的確にとらえているといえる.

また, 被験者による判定と比較するため, 「MS-Level による重要度の自動判定」も行った. 「重要度の自動判定」とは, 該当シーンにおける会議参加者全員の MS-Level の平均値 avg に基づいて「重要 ($1 \geq avg \geq 0.8$)」, 「やや重要 ($0.8 > avg \geq 0.6$)」, 「普通 ($0.6 > avg \geq 0.4$)」, 「やや不要 ($0.4 > avg \geq 0.2$)」, 「不要 ($0.2 > avg \geq 0$)」のように判定したものである.

表 4 を見ると, 方法 1 よりも方法 2, 方法 2 よりも方法 3 の方が, マッチング率が高いことが分かる. つまり, 方法 1~3 の中では, 動画メディアと MS-Index を利用した方法 3 が一番的確な判定が行えるという結果になった. この差が有意なものであることを確認するために, 方法 1 と方法 2, および, 方法 2 と方法 3, 方法 1 と方法 3 の間で Wilcoxon 符号順位和検定を行ったところ, 棄却率はそれぞれ $p = 0.0517, 0.00420, 0.00280$ ($N = 24$) となった. つまり, 方法 1 と方法 2 の間には有意水準 5% で有意に差が見られないが, 方法 2 と方法 3, および, 方法 1 と方法 3 の間にはそれぞれ有意水準 1% で有意に差がある. よって, 会議を撮影した動画メディアの重要

度を判定する際には、動画メディアと MS-Index の両方を利用した方が、よりの確かな判定が行えるといえる。

また、創造会議、決定会議、調整会議の3種類の動画メディアに対して、それぞれ MS-Level で自動判定を行ったところ、マッチング率は平均 0.433 となり、方法 1-3 のどの場合よりも劣った結果となった。これは、現実の会議には様々な要因が含まれており、判定には人間の高度な判断能力が必要であるためと思われる。さらに、重要度の自動判定のアルゴリズム（会議参加者全員の MS-Level の平均値のみからシーン重要度を判定）が非常に単純すぎた点も問題である。しかし、自動判定が実用的なものとなれば、各シーンの重要度判定や議事録起こしに人間が費やしていた時間や手間を大幅に軽減することができるので、今後はアルゴリズムなどを再検討し、実現に向けて研究を進めていく予定である。

8. 結 論

本稿では、会議を撮影した動画メディアの思考状態インデキシングを提案した。思考状態の測定には思考と関連が深い β 波を利用し、MS-Level という形式で数値化した。そして、MS-Level をもとに MS-Index という思考状態を表す索引を作成し、会議を撮影した動画メディアと組み合わせて、会議の各シーンにおける各参加者の思考状態が把握できるようなプロトタイプを作成した。プロトタイプを利用した実験では、MS-Level が思考状態を的確にとらえていること、そして、MS-Index が思考状態の分析に有用であることが確認できた。

動画メディアに思考状態インデキシングを施せば、いつ、誰が、どのような思考状態にあったか把握することが容易になり、そのシーンを簡単に参照することも可能である。たとえば、会議において自分の意見に強い関心を示していた人物を知る手がかりにしたり、商談において顧客が関心を示した話題を抽出したりする利用法が考えられる。

また、教育の場や演説の場に導入して演説者の発言に対する評価を聴衆の思考状態から自動集計するといった利用法も考えられる。これは、3.4 節で述べた「場の雰囲気」という概念を利用したものであり、今後研究を進めていくべきであると考えている。

そのほかにも、映画やスポーツの観客全体の思考状態を集計することで、映画・スポーツの「ダイジェスト自動生成」を行うことも可能と思われる。ダイジェスト生成に関しては、各個人ごとに自分の思考状態だけを反映した「個人的なダイジェスト」を生成するこ

とも有益であると思われる。

参 考 文 献

- 1) Chiu, P., Kapuskar, A., Reitmeier, S. and Wilcox, L.: Room with a Rear View: Meeting Capture in a Multimedia Conference Room, *IEEE MultiMedia*, Vol.7, No.4, pp.48-54 (2000).
- 2) Yong, R., Gupta, A. and Cadiz, J.: Viewing meetings captured by an omni-directional camera, *Proc. SIGCHI Conference on Human factors in computing systems*, pp.450-457 (2001).
- 3) Waibel, A., Bett, M., Finke, M. and Stiefelhagen, R.: Meeting Browser: Tracking and Summarizing Meetings, *Proc. Broadcast News Transcription and Understanding Workshop*, pp.281-286 (1998).
- 4) Hindus, D., Schmandt, C. and Horner, C.: Capturing, Structuring, and Representing Ubiquitous Audio, *ACM Trans. Inf. Syst.*, Vol.11, No.4, pp.376-400 (1993).
- 5) 下田 宏ほか：動的顔画像からのリアルタイム表情認識システムの試作、ヒューマンインタフェース学会論文誌, Vol.1, No.2, pp.25-32 (1999).
- 6) 本田新九郎, 富岡展也, 木村尚亮, 大澤隆治, 岡田謙一, 松下 温: 作業者の集中度に応じた在宅勤務環境の提供—仮想オフィスシステム Valentine, *情報処理学会論文誌*, Vol.39, No.5, pp.1472-1483 (1998).
- 7) Picard, R.: *Affective Computing*, The MIT Press (1997).
- 8) Giannitrapani, D.: The Role of 13-Hz Activity in Mentation, *The EEG of Mental Activities*, pp.149-152 (1988).
- 9) 加藤象次郎ほか：初学者のための生体機能の測り方, 日本出版サービス (1999).
- 10) 本田研作, 福井健太郎, 宮田章裕, 重野 寛, 岡田謙一: 思考状態アウェアネスによる対人コミュニケーション支援, マルチメディア, 分散, 協調とモバイル (DICOMO 2003) シンポジウム, pp.449-452 (2003).
- 11) 福井健太郎, 本田研作, 宮田章裕, 重野 寛, 岡田謙一: 多地点遠隔コミュニケーションにおける思考状態アウェアネスの提供, 日本バーチャルリアリティ学会研究報告 CSVC2003-4, pp.33-38 (2003).
- 12) 本田研作, 福井健太郎, 岡田謙一: 脳波情報を利用したアバタの表情変化, 日本バーチャルリアリティ学会研究報告 CSVC2002-14, pp.19-24 (2002).
- 13) 宮田章裕, 福井健太郎, 本田研作, 重野 寛, 岡田謙一: 脳波情報を利用したミーティング分析, 第113回マルチメディア通信と分散処理研究会 2003-DPS-113, pp.63-68 (2003).

- 14) 相澤清晴, 石島健一郎, 椎名 誠: ウェアラブルによる体験映像の取得と脳波による要約の自動生成, 電子情報通信学会論文誌 D-2, Vol.J86-D-2, No.6, pp.807-815 (2003).
- 15) 中村亮太, 市村 哲, 松下 温: スポーツ中継の知識を用いた多視点リプレイ自動生成システム, 第49回グループウェアとネットワークサービス研究会 2003-GN-49, pp.127-132 (2003).
- 16) 武者利光: 「こころ」を測る, 日経サイエンス 1996年4月号, pp.20-29 (1996).
- 17) Sakai, K., Hikosaka, O., Miyauchi, S., Takino, R., Sasaki, Y. and Putz, B.: Transition of brain activation from frontal to parietal areas in visuomotor sequence learning, *Journal of Neuroscience*, pp.1827-1840 (1998).
- 18) 高橋 誠: 会議の進め方, 日本経済新聞社 (1987).

(平成 15 年 11 月 12 日受付)

(平成 16 年 9 月 3 日採録)



宮田 章裕 (学生会員)

2003年慶應義塾大学理工学部情報工学科卒業。現在, 同大学大学院理工学研究科開放環境科学専攻修士課程に在学中。グループウェア等の研究に従事。2004年 DICOMO シンポジウム最優秀プレゼンテーション賞受賞。



福井健太郎 (正会員)

2001年慶應義塾大学理工学部情報工学科卒業。2003年同大学大学院理工学研究科開放環境科学専攻修士課程修了。現在, 同大学大学院理工学研究科博士課程に在学中。グループウェア等の研究に従事。2004年 IEEE Computer Society Best Paper Award 受賞。



本田 研作

2004年慶應義塾大学大学院理工学研究科開放環境科学専攻修了。



重野 寛 (正会員)

1990年慶應義塾大学理工学部計測工学科卒業。1997年同大学大学院理工学研究科博士課程修了。1998年同大学理工学部情報工学科助手(有期)。現在, 同大学理工学部情報工学科助教授。工学博士。計算機ネットワーク・プロトコル, モバイル・コンピューティング, マルチメディア・アプリケーション等の研究に従事。著書『~ネットワーク・ユーザのための~無線LAN技術講座』(ソフト・リサーチ・センター)、『コンピュータネットワーク』(オーム社)等。電子情報通信学会, IEEE, ACM各会員。



岡田 謙一 (フェロー)

慶應義塾大学理工学部情報工学科教授, 工学博士。専門は, CSCW, グループウェア, コンピュータ・ヒューマン・インタラクション。『ヒューマンコンピュータインタラクション』(オーム社)、『コラボレーションとコミュニケーション』(共立出版)をはじめ著書多数。情報処理学会誌編集主査, 論文誌編集主査, GW研究会主査等を歴任。現在, 情報処理学会 GN研究会運営委員, BCC研究グループ幹事, 日本 VR学会仮想都市研究会副委員長。IEEE, ACM, 電子情報通信学会, 人工知能学会各会員。1995年度情報処理学会論文賞, 情報処理学会40周年記念論文賞, 2000年度情報処理学会論文賞受賞。