

マルチモーダル情報による問題解決型会議における肯定/否定的発話の認識方式の検討

A Study on Recognition Method of Positive / Negative Utterance in Problem Solving Meeting with Multimodal Information

上原 佑太郎

Yutaro Uehara

岡山大学 阿部研究室

Abe Laboratory, Okayama University

概要 本報告では、会議における人間の肯定的、否定的発話の認識のための「会議音声」と「会議映像」の収録方式を述べる。会議形態や参加者へのインストラクション、収録環境に着目する。本報告では、参加しやすい議題で論理的な議論を行い、自然な会議を目指す。また、そのような会議における音声や表情、ジェスチャの収録の検討を行った。

1 はじめに

会議とは、人間同士の対話によってさらに特定の議題に対して何等かの解を得る方法としてよく行われており、その種類としては、主にディスカッション (Discussion) とディベート (Debate) がある。前者はある議題に対して、各々が自身の意見を言い合う議論である。一方、後者はある議題に対して賛成 (Positive) か反対 (Negative) に分かれた2つのチームが一定のルールに従って行う議論である。さらに、ディスカッションは大きく自由討論型、選択型、問題解決型の3つに分けられ、それぞれ議題について自由に話し合う、複数の選択肢から1つを選び出したり順位をつける、ある事例に対する課題の解決法を議論により導く、といった議論である。

また、会議は長時間にわたる場合が多く、見返すのに非常に手間がかかる。そのため、誰がどのような論点で議論していたのかを把握するために見返すのは現実的ではない。そのような会議情報を把握する為の会議の分析は様々なアプローチで行われている。話者認識の分野では、会議音声を分析窓 (Analysis Window) を用いて話者の分割を行い、会議中の話者が誰なのかを認識している。[1] また、言語情報処理の分野では、会議のテキスト情報 (議事録) からグラフベースのフレームワークを用いて会議要約が行われている。[2]

ある特定の議題または意見に対しての肯定や否定の認識によりアノテーションをつけることができれば、会議音声等の会議情報の素早い理解に有用であると考えられる。そこで、本報告では会議における肯定や否定の認識のための音声や映像の収録方式を述べる。

2 対象とする会議形態

ここでは、どのような会議形態を収録するかについて [3][4] を参考に述べる。本研究では、次の三点を重視する。これら3点を満たす会議形態として、問題解決型会議を扱うこととした。

1. 参加しやすい議題であること
2. 議論が論理的であること
3. 肯定、否定の認識に適した会議進行であること

一つ目の参加しやすい議題を扱いたい理由は、ジェスチャ等のパラ言語情報や音声の言語情報といった参加者全員の積極的な発言、振る舞いを撮影、録画したいのである。例えば、「スネアドラムの品質を向上させるにはどうすれば良いか」といったメーカーにしか深く議論できない議題よりも「満員電車を減らすためにはどうすればよいか」という馴染みのある事柄について議論の方が議論は白熱すると考えられる。

二つ目の、議論が論理的であることの理由としては、肯定とも否定とも取れない発言を避けるためである。それは、肯定と否定の認識の観点で妨げとなる。また、議論が論理的でない場合、議論の道筋を横切って感情論で押し通そうとする参加者が予想されるため、会議とは言えなくなる。

最後に、三つ目の肯定、否定の認識に適した会議進行が必要となる理由としては、例えば、ディベートの進行は肯定派と否定派とではっきりと分けられ、フレームワークとして定まっているため本研究には不適切であるということである。

3 会議音声・映像の収録方式

ここでは、前章 (2) で決定した問題解決型会議の収録方式について述べる。

3.1 参加者

参加人数は参加者が発言しやすく、短時間である程度深く議論ができるとされる4, 5人とする。積極的な発言やジェスチャ、表情の移り変わり、韻律の変化の収録を目的とするため、ある程度お互いを知っており、親密度が高いメンバを募る。

3.2 参加者へのインストラクション

会議参加者へのインストラクションは簡潔であることが好ましく、長々としてしまうと返って混乱を招く。そのため、問題解決型会議の進め方をインストラクションとしては伝えず、ファシリテータが進め方に沿って議論を導くか、伝えてから、参加者の議論の進め方にファシリテータが合いの手を入れるという2通りが考えられる。これらの欠点としては、前者はファシリテータをする人のスキルが未熟な場合、議論に割って入る可能性があり、議論の妨げとなるということである。また、後者は参加者が決まったフレームワークの基で議論を進めるので発言しにくさが生まれる可能性があるということが挙げられる。この問題に対しては、テスト収録後のアンケートの結果を参考に決定する。

基本的な参加者へのインストラクションは、全員同時に収録直前に行う。そこでの説明の概要を以下に示す。

1. 今日収録させていただくデータは、音声と表情、ジェスチャなどの研究用に収録させていただき、その目的以外に利用しない。
2. 目的は、会議での私たち人間の会話における肯定や否定の分析にあり、実際に起こりえる現象は、そのまま収録したい。
3. きれいな声や話し方等の意図的な劇を望んでいるわけではなく、データ収録であることはなるべく意識せずに、会議を行っていただきたい。

次に論理的な議論をするための注意点を三点以下に示す。

1. 目的と手段、原因と結果、根拠と結論をつなげる
2. 賛成とも反対とも取れないような意見はなるべく避ける
3. 論点を共通認識として念頭に置く

これらは議論を円滑に進める為の一般的な注意点である。しかし、このまま伝えるのでは抽象的である為、参加者にとってイメージしにくい。そこで、注意点のインストラクションでは具体例を交えながら説明する。

会議の時間は1時間程度を想定する。また、補足として、補助道具でメモは持参してよいこととする。

3.3 収録環境

ここでは、予定している収録環境を述べる。参加者が4人の場合の収録環境を図1に示す。音声の収録は、参加者全員にピンマイクを装着し、更に机の横に全体の音声を収録するマイクを1台設置する。映像の収録には、机中央のカメラ2台で参加者の表情、ジェスチャが映るようにし、全体が映るようにカメラ3を1台設置する。カメラ1、カメラ2に関してはWeb会議用の広角レンズを搭載したものをを用いる。

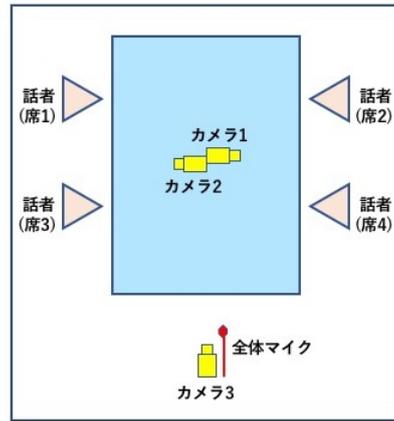


図1: 収録環境

3.4 議題の例

ここでは想定している議題の例をいくつか挙げる。

- 大学生の抱える問題を一つ挙げ、解決策を考えよ
- ただの石ころを3万円で売る方法
- シニア世代にIT機器を普及させる方法
- 上司とうまくコミュニケーションをとるには

4 まとめ

本報告では、発話態度の認識のための会議音声・映像の収録方式について述べた。会議には目的に応じて様々な会議形態があり、発話態度の認識には問題解決型会議が最適である。また、収録環境や参加者へのインストラクションについても述べた。収録環境は、参加者の発言、表情やジェスチャが収録できるような設置を想定したが、テスト収録によっては改善の可能性はある。インストラクションについても議論の進め方を参加者に伝えるか否かはテスト収録によるので課題は残る。今後の予定としては、テスト収録である。また、収録した後、収録方式を調整し発話態度の認識に関する研究の調査が挙げられる。

参考文献

- [1] Leonardo V. Neri, Hector N. B. Pinheiro, Tsang Ing Ren, George D. da C. Cavalcanti, André G. Adami "SPEAKER SEGMENTATION USING I-VECTOR IN MEETINGS DOMAIN," in *Proc. ICASSP*, pp. 5455–5459, Mar, 2017.
- [2] Mohammad Hadi Bokaei, Hossein Sameti, Yang Liu "Extractive Meeting Summarization through speaker zone detection," in *Proc. INTERSPEECH*, pp. 2724–2728, Sep, 2015.
- [3] 堀 公俊, 加藤 彰 (2018). "ロジカル・ディスカッション," 日本経済新聞出版社.
- [4] 堀 公俊, 加藤 彰 (2017). "ワークショップ・デザイン," 日本経済新聞出版社.