

レジュメ型オンライン会議における 音声の言語情報を用いた スクローリングインタラクションの基礎検討

森本 くるみ^{1,a)} 松森 匠哉¹ 奥岡 耕平¹ 木本 充彦¹ 今井 倫太¹

概要: 近年テレワークを導入する組織が増加傾向にあり、レジュメを画面共有してオンライン会議を進行するシーンが多く見られるようになった。レジュメ型オンライン会議では、司会進行とレジュメの画面共有を異なる参加者が担う場合がある。例えば、司会者がモバイル端末から参加しており画面共有できない時である。司会者と画面共有者が異なる場合、画面共有者は、司会者の発話内容を汲み取ってレジュメを適切にスクローリングする必要がある。一方司会者は、スクローリング動作が意図通りに行われた場合は会議を続行し、意図通りでなかった場合にはスクローリング命令を発する必要がある。この司会者と画面共有者の意図の読み合いを、我々はスクローリングインタラクションと呼ぶ。スクローリングインタラクションが成功すれば、会議を円滑に進めることができる。そこで本研究では、オンライン会議の録画データから、スクローリングインタラクションが発生する要因を分析する。分析を基に、司会者の発話内容から自動でスクローリングを行うシステムのプロトタイプを作成し、言語によるスクローリングインタラクションの基礎検討を行う。

1. 序論

近年テレワークを導入する組織が増加傾向にあり、レジュメを画面共有してオンライン会議を進行する、レジュメ型オンライン会議が多く見られるようになった。レジュメ型オンライン会議では、司会者がモバイル端末から参加しており画面共有できない時など、司会者と画面共有者が異なる場合がある。司会者と画面共有者が異なる場合、画面共有者は、司会者の発話内容と関係している箇所を表示するように、レジュメを適切にスクローリングする必要がある。一方司会者は、適切な箇所が表示されている場合は会議を続行し、話題と関係ない箇所が表示されている場合には、画面共有者にスクローリングを促す必要がある。この司会者と画面共有者の意図の読み合いを、我々はスクローリングインタラクションと呼ぶ。スクローリングインタラクションが成功すれば、話題と画面共有されたレジュメの間に対応関係ができるため、より円滑にオンライン会議を進行できる。そこで本研究では、画面共有者の役割を担い自動でスクローリングを行うシステムのプロトタイプを作成する(図1)。そしてレジュメ型オンライン会議にお

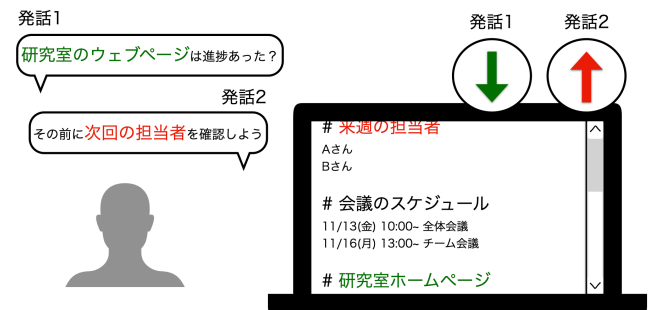


図1 司会者の発話内容に基づいた自動スクローリングシステム

ける、音声の言語情報を用いたスクローリングインタラクションの基礎検討を行う。

スクローリングインタラクションにおいて、画面共有者は司会者の発話とレジュメの内容を踏まえて、話題に関係する箇所を画面に表示することが課題となる。画面共有者の役割を担うシステムは、司会者の話題とレジュメの内容を理解し、適切な方向と移動量を決定してスクローリングを実行する必要がある。

先行研究では、コンピュータの音声ユーザーインターフェースとして、SUITEKey システム [1] が開発されている。SUITEKey では音声によるキーボード入力や、マウスカーソルの移動およびクリックが実行できる。マウスカー

¹ 慶應義塾大学
Keio University

^{a)} morimoto@ailab.ics.keio.ac.jp

ソルを下に移動させる場合は、「マウスを下に移動」と発声するとカーソルが動き始め、「停止」と発声すると止まる。また、「マウスを3つ下げる」や「マウスの位置を(3,5)に設定」のように、相対位置や絶対位置を用いて移動させることも可能である。五十嵐らは音声のもつ音響信号情報に着目し、ユーザーの発声から結果のフィードバックまでに時間のかからない、連続的に機械操作できるインターフェース [2] を提案した。筆者らの開発したインターフェースでは、「上へ移動、あーあー」のように、発声をしている間スクローリングを実行し続けることが可能である。

既存の研究には2つの問題点がある。1つは「上へ移動」のようにスクローリングのトリガーとなる特定のキーワードを発声しなければならない点である。もう1つはシステムの出力内容を一切考慮していない点である。出力内容を考慮せず音声のみを参考しているため、指定された座標への移動や「停止」と言うまで、もしくは発声している間スクローリングを続けるシステムになっている。

本研究では音声の言語情報およびシステムに出力されたレジュメの内容を加味し、話題に関係している箇所までレジュメをスクローリングするシステムを提案する。

本稿では2章で関連研究を述べる。3章でオンライン会議の録画データによるスクローリングインタラクションの要因分析を行う。3章の分析結果を踏まえ、4章で自動でスクローリングを行うシステムのプロトタイプを提案し、5章でスクローリングシステムを用いた実験内容および実験結果を示す。6章で実験結果の考察、7章で今後の展望を示し、最後に8章で結論を述べる。

2. 関連研究

近年では、特定のキーワードだけでなく操作に関する類義語を認識できるゲーム操作システム [3] が吉田らによって開発された。Unity の音声認識エンジンおよび日本語概念辞書を用いて、操作に関する単語に反応しゲームが実行できる。特定のキーワード以外にも反応できるシステムである点が本研究の目的と類似している。しかし音声の言語情報のみに着目しており、システムの出力であるゲームの操作画面の情報が一切用いられていないため、限られた5つの操作の実行に留まっている。

一方福地らは、音声を用いず指のストローク動作のみに着目したスクローリングインタラクションシステム [4] を開発し、ケーススタディに成功した。指のストローク動作および画面の構成要素からスクローリング量を予測し、興味のある箇所までのスクローリングおよび興味のない箇所のスキップを実現している。音声情報を一切用いていないため本研究の目的とは異なっているが、システムの出力画面の構成要素を加味してスクローリング量を決定している点が類似している。

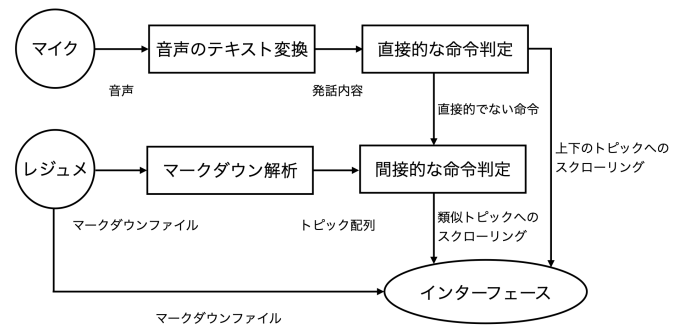


図2 システム構成

3. スクローリングの要因分析

スクローリングシステムを開発するにあたり、オンライン会議の録画データから、スクローリングインタラクションが発生する要因を分析した。6本のオンライン会議録画データにおける合計80回のスクローリングについて、スクローリングが発生する要因となった司会者の発話を文字起こし記録した。オンライン会議は同一の司会者が進行する研究室内のミーティングであり、2名の画面共有者がそれぞれ3本ずつスクローリングを担当した。

ビデオ分析の結果、司会者の発話を直接的なスクローリング命令と間接的なスクローリング命令の2種類に分類できることが明らかになった。直接的なスクローリング命令は、「下行って」、「戻って」、「続きまして」のように、方向や遷移を意味するキーワードによってスクローリングを促す発話を指す。間接的なスクローリング命令は、直接的な命令を除いたスクローリングの要因となる発話である。「〇〇の時間は」、「〇〇は進捗あった?」のように、レジュメに書かれたトピックに関係する発話によって、該当の箇所までスクローリングを促すものが当てはまる。

4. 提案システム

本論文で提案するシステムは、司会者の発話とレジュメの内容を認識し、直接的な命令と間接的な命令に反応してスクローリングを実行する。特定のキーワードを含む直接的な命令だけでなく、レジュメとの関係性に基づいた間接的な命令も扱うことで、より自然で円滑な会議の進行が可能になる。

システム構成を図2に示す。入力はマイクからの音声およびレジュメファイル、出力はスクローリングである。

4.1 音声テキスト変換

マイクから入力された音声は、Google社のCloud Speech-to-Text APIによってテキストに変換される。

4.2 マークダウン解析

システムはマークダウン形式のレジュメを認識し画面に

表示する。起動時にマークダウンの構文解析を行い、H1 タイトルの文字列と座標をレジメのトピック配列として保持する。

システムは常に現在話題になっているトピックを、トピック配列のインデックスとして記録する。システム起動時のインデックスは0であり、スクローリングが実行されるたびに値は上書きされる。

4.3 直接的な命令判定

テキスト変換された発話内容は、以下のキーワードを含む場合は直接的な命令と判定される。

スクローリングのキーワード

- 上へのスクローリング
前, 戻, 上
- 下へのスクローリング
次, 続, 下

「下行って」や「続きまして」は、それぞれ「下」「続」のキーワードを含んでいるため、直接的な命令としてスクローリングのトリガーになる。上にスクローリングを実行する直接的な命令と判断されれば、現在の話題の1つ前のトピックタイトルの座標まで、下にスクローリングする命令と判断されれば、1つ後ろのトピックタイトルまでスクローリングする。

4.4 間接的な命令判定

直接的な命令でないと判定された発話内容は、間接的なスクローリング命令の判定に移る。入力発話内容、現在の話題から前後5つのトピックである。命令判定モジュールは、受け取った文字列から形態素解析によって名詞のみを抽出し、Word2Vec モデルでベクトル化した後、発話内容とトピックそれぞれに含まれる名詞ベクトル間のコサイン類似度を計算する。コサイン類似度が閾値を超えた場合は、類似トピックまでスクローリングが実行される。Word2Vec モデルは日本語 Wikipedia を用いて、独自に学習を行った。

5. 実験

5.1 実験方法

3人の異なる司会者によるそれぞれ約1時間の研究室内のミーティング映像とレジメを用いて、システムが意図通りのスクローリングを実行するか検証した。映像の中からスクローリングのトリガーとなった司会者の発話のみを抽出し、マイクに発声することで入力した。

抽出したスクローリング命令によって、司会者が意図したトピックタイトルまでスクローリングされた場合に成功とみなした。スクローリング命令は連続的に発せられるため、1つ誤ったスクローリングを実行した場合、以降の命令が全て失敗してしまう恐れがある。そこで本実験では、ス

表 1 3 会議における成功・失敗したスクローリング命令数の内訳

		会議 A	会議 B	会議 C	合計
直接	成功	18	12	6	36
	失敗	3	0	0	3
間接	成功	12	11	14	37
	失敗	11	14	24	49
合計		44	37	44	125

スクローリングに失敗した場合は次のスクローリング命令を入力する前に、命令が発話される直前の座標までスクローリングすることとした。

5.2 実験結果

3つの会議における、成功・失敗したスクローリング命令数の内訳を表1に示す。合計125のスクローリング命令のうち、直接的な命令は39、間接的な命令は86であった。125の命令のうち、スクローリングに成功した命令は73、失敗した命令は52であった。直接的な命令は39命令のうち36命令成功、3命令が失敗、間接的な命令は86命令のうち37命令が成功、49命令が失敗であった。

6. 考察

6.1 成功した直接的な命令

「次」、「下」というキーワードに反応して、現在の話題から次のトピックへのスクローリングに成功した。

具体例

- 「で次はえっと」
- 「下に行って」
- 「えっと下に行っていただいていたいいですか」
- 「次に移りたいと思います」

6.2 失敗した直接的な命令

本システムは「そして」という単語に反応して、次のトピックにスクローリングすることができなかった。人間の場合は「そして」、「それから」、「あとは」などの単語に反応してスクローリングを実行できるが、本システムはキーワードに指定されていない単語から、遷移の意図を読み取ることはできない。より多くの単語を直接的なスクローリング命令のキーワードとして追加する方法もあるが、意図しないスクローリングが増えるというトレードオフがある。

6.3 成功した間接的な命令

間接的な命令がスクローリングに成功する例は2種類ある。1つはトピックタイトルに含まれている名詞を司会者が発話した時である。この場合は、Word2Vec で同じ単語同士を比較することになるため、類似度は1.0となり、該当するトピックまでスクローリングが実行される。もう1つは名詞の比較が閾値を超え、類似していると判断された場合である。類似した例には、「Birthday」と「バースデー」、

「Web」と「ウェブ」などが見られた。

具体例

- 「〇〇のレイアウト考えなきゃな」
トピック「〇〇 レイアウト」まで移動に成功
- 「B3 向けのものを」
トピック「B3 ウェブ」まで移動に成功
- 「ウェブページを〇〇に移行したい」
トピック「研究室の Web」まで移動に成功
- 「〇〇先生のバースデーパーティー」
トピック「〇〇 Birthday」まで移動に成功

6.4 失敗した間接的な命令

失敗した間接的な命令の具体例と解決策を原因によって分類し以下に示す。

6.4.1 意図しないトピックに反応した例

「ネットワーク」と「リンク」、「論文」と「レポート」などが関係していると判断され、意図せず別の箇所にスクローリングされた。閾値の調整や、命令判定のアルゴリズムの見直しによって改善が見込まれる。現在の座標からのスクローリング距離とコサイン類似度の両者を加味することで、Word2Vec で意図しないトピックに反応することのある程度防ぐことができると考えられる。

具体例

- 「サーベイレポート」
トピック「研究室論文情報 英語版」まで移動し失敗
- 「題目申請期間」
トピック「残留許可願い」まで移動し失敗
- 「M1 スケジュール」
トピック「B3 研究室紹介」まで移動し失敗

6.4.2 英単語と日本語の関係が取れなかった例

「学会」と「Conference」のように、英単語と日本語の関係が取れず、スクローリングに失敗した例もあった。Word2Vec モデルの改善や別のモデルの使用が必要である。

6.4.3 未知の単語をベクトル化できなかった例

ジェネミ (ジェネラルミーティングの略) のように、独自で利用しているような略語に反応することができなかった。ベクトル化できない単語については、文字列同士での比較が必要と考えられる。

6.4.4 トピックタイトルから連想できない発話がされた例

トピックタイトルではなく、タイトルに紐づいた本文に含まれるキーワードに言及した際など、タイトルの内容のみからは連想できない発話によって失敗する例も見られた。タイトルに紐づく本文も加味した文脈で比較するよう改善する必要がある。

6.4.5 誤って直接的な命令と判断された例

直接的なスクローリング命令と誤って判定されて、意図しないスクローリングが実行されたものもあった。「〇〇は以下のようなです」などがこの失敗例に該当する。現在は

直接的な命令と間接的な命令を独立して評価しているが、直接的なキーワードを含む発話に関しても、Word2Vec で評価をした上でスクローリングを決定する必要があるだろう。

7. 今後の展望

本論文ではオンライン会議の中からスクローリングのトリガーとなった発話のみを抽出し実験を行ったが、抽出を行わず全ての発話を入力してシステムの有効性を検証する必要がある。

また、本論文では言語的なスクローリング命令のみを扱ったが、人間の場合は暗黙的に行間を読んだスクローリングを実行することがあった。話の終わりや行間を読み取り、システムが自発的にスクローリングを行うことも必要だろう。

8. 結論

本論文では、レジュメ型オンライン会議における司会者と画面共有者の、音声の言語情報を用いたスクローリングインタラクションを検討した。まずオンライン会議の録画データを分析し、司会者の発話を、方向や遷移を表すキーワードを含む直接的なスクローリング命令と、レジュメに関係した発話から成る間接的なスクローリング命令に分類した。そして、2種類の命令とレジュメの内容を理解し、スクローリングを実行をするシステムのプロトタイプを開発した。開発したシステムに、3人の異なる司会者によって発せられた125命令を流し込み、スクローリングが意図通りに行われるか実験を行った。「次は」、「下に行って」のような直接的な命令39命令のうち36命令のスクローリングに成功するだけでなく、「先生のバースデーパーティー」のような間接的な命令86命令のうち37命令のスクローリングに成功した。今後は間接的な命令の認識精度を向上させるために、判定アルゴリズムの改善に取り組んでいく。

謝辞 本研究は、JST, CREST, JPMJCR19A1, および JSPS 科研費 JP20K19897 の支援を受けたものである。

参考文献

- [1] Manaris, B., McCauley, R. and MacGyvers, V.: An intelligent interface for keyboard and mouse control, Proc. 14th Int'l Florida AI Research Symposium (FLAIRS-01), Citeseer, pp. 182-188 (2001).
- [2] 五十嵐健夫ほか: 言語情報を用いない音声による直接操作インタフェース, 情報処理学会研究報告ヒューマンコンピュータインタラクション (HCI), Vol. 2004, No. 74 (2004-HI-109), pp. 47-51 (2004).
- [3] 吉田美乃里, 神部真音, 備前比呂, 川合康央ほか: 日本語音声認識を用いたキャラクターの操作システム, エンタテインメントコンピューティングシンポジウム 2020 論文集, Vol. 2020, pp. 150-153 (2020).
- [4] Fukuchi, Y., Takimoto, Y. and Imai, M.: Adaptive enhancement of swipe manipulations on touch screens

with content-awareness, *12th International Conference on Agents and Artificial Intelligence, ICAART 2020*, SciTePress, pp. 429–436 (2020).