

## マルチストリーミングの音声情報による画面調整機能 Voice Based Layout Adjustment Function for Multi-Streaming

齋藤 溪<sup>†</sup> 橋本浩二<sup>†</sup>

<sup>†</sup>岩手県立大学ソフトウェア情報学部

### 1. はじめに

情報技術の発展によりインターネットの端末同士でも映像や音声を双方向で通信することが容易に可能となった。その通信システムとして、テレビ会議システムや Web 会議システム、ビデオチャットなどが幅広く利用されており、遠隔地における複数地点の通信や複数人の通信も可能となっている。

一方、複数地点からの映像を扱うことを想定すると、各地点における映像表示画面の調整方法には検討の余地がある。例えば遠隔会議中、話者がいる地点の映像を適宜切り替えて表示したり、複数地点間で議論するような場合における画面を適切に調整したりすることは、決して容易ではない。結果として、視聴者から見て話者を特定できない状況や、議論中の参加者の一部しか表示されていない状況が発生する。これは、既存システムを運用する際の、カメラスイッチやレイアウト変更操作における課題となっている。

画面調整を行うためにはビデオスイッチャーのように映像を切り替える機器がある。しかしその切替操作は自動化に至っておらず、Web 会議システムやビデオチャットでカメラスイッチを行うには通常、操作者を必要とする。これまでに、カメラスイッチの自動化のためにマイクの配置の仕方と音声の大きさに基づく画面調整を行う研究[1]や、視線に基づく画面調整を行う研究[2]が行われている。これら既存の研究における手法は、複数の参加者の相対的な位置が固定されていることを想定している。一方、持ち運び可能な端末を用いて、場所を問わず会議を行う場合などを考慮すると、音声の大きさのみで画面を調整したりマイクの配置を予め調整したりすることは困難であり、新しい画面調整方法が必要になると考えられる。

以上のことを踏まえて様々な会議用途に対応できる画面調整可能な機能を提案する。

### 2. システム概要

本研究では音声に着目し、その会話の内容も考慮した画面調整機能の実現を目指している。例えば、発言者が参加者の名前を呼んで対話が始まる場合が多くある。その名前の発言を検出することにより発言者の特定と話相手の特定を行うことができると考えられる。図1は本研究におけるシステム概要図である。現状では配信映像もとの Stream Source に対し Staff が手動で Display の画面調整を行っているが、本研究では Streaming Terminal 上で自動的に画面調整を行う。画面切り替えは発言者が話相手を特定するための機能、レイアウト変更は対話している人がわかるような表示に変更する機能、拡大機能は話者を特定するためにレイアウトに反映させる機能であり、これらの機能を組み合わせて画面調整を行う。

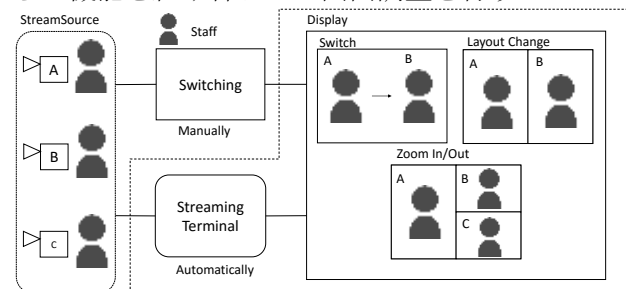


図1: 提案システム概要

図2ではシステムアーキテクチャを示している。会話を認識するために音声認識モジュールを VoiceRecognizer と音声認識のイベントを検出する VoiceHandler で構成している。ViewController は音声の認識結果に応じて、ストリームの表示を行っている VideoElement の操作と画面のレイアウト変更を行う LayoutManager に画面調整のメッセージを送信する。LayoutManager は Viewer の画面構成を変更するためのレイアウト情報を扱う。VideoElement はストリーム表示のため、ウィンドウ大きさと、ユーザ名と、ストリームの情報を持つ。Streamin Module は VideoElement にストリームの情報とそれぞれのクライアントが通信を行うためにシグナリング情報を Singnalling Server から受け取る。それぞれのモジュールを Application として配置するための Application Server でシステムを構成する。

Voice Based Layout Adjustment Fucntion for Multi-Streaming  
Kei Saito<sup>†</sup> and Koji Hashimoto<sup>†</sup>  
<sup>†</sup>Faculty of Software and Infomation Science, Iwate Prefectural University

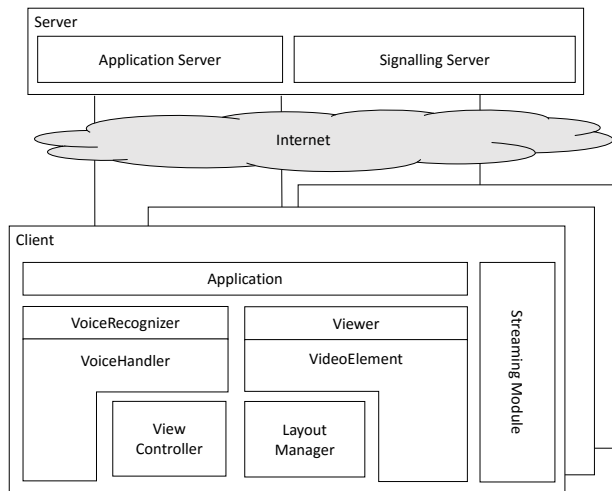


図2: システムアーキテクチャ

### 3. 音声情報による画面調整

図3はシステム機能を示している。以下に機能の詳細を示す。

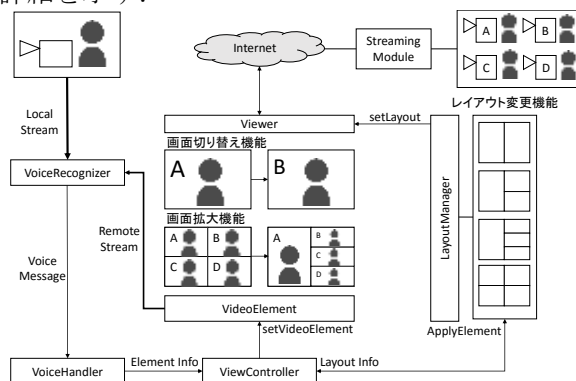


図3: システム機能

#### (1) 画面切替機能

画面切り替え機能では発言者の名前とその応答を検出し、応答があったユーザの画面を切り替えて表示を行う。View Controller は VoiceHandler から受け取った Element Info を基に VideoElement を表示させる。この際 Element Info は発言者と名前の呼ばれた話相手の VideoElement の情報を持つ。この画面切り替え機能によって発言者は話相手を特定することが可能となる。

#### (2) 画面拡大機能

画面拡大機能では発言者を音声の認識をトリガーとして拡大表示する。View Controller は VoiceHandler から受け取った Element Info を基に VideoElement を拡大表示のレイアウトのに配置する。この際 Element Info は音声認識の開始のトリガーがある VideoElement の情報を持つ。この画面拡大機能によって発言者の特定を可能とする。

#### (3) レイアウト変更機能

レイアウト変更機能では Viewer の画面構成を音声情報に基づきレイアウト情報を提供する。

LayoutManager は ViewController から受け取った ApplyElement を基にレイアウトを構成する。このレイアウト変更機能によって対話の方向性の伝達と、参加者が多い場合でも発言者の特定が可能な画面調整を行うことができる。

### 4. プロトタイプシステム

システムアーキテクチャの Client は GoogleChrome を利用し、音声認識を行う VoiceRecognizer は WebSpeechAPI。その音声認識の結果を処理する VoiceHandler は Javascript, Viewer と VideoElement は HTML5, ViewController は jQuery, LayoutManager は CSS, Streaming Module は WebRTC と Socket.io, Application Server は GitHubPages, Signalling Server は Firebase で実装を行った。

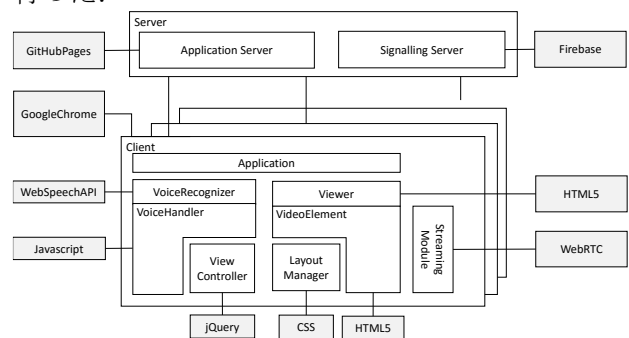


図4: プロトタイプシステム

### 5. まとめと今後の課題

本稿では、音声情報による画面調整機能について提案した。会話の内容を含める音声情報の画面調整を行うアプローチにより発言者の検出とその話相手を検出することが可能となった。一方で、画面調整後に他のユーザが音声の入力を行うと発言者にとって適切な画面表示を行えない場合もある。画面調整後どのタイミングで再び画面調整を行うべきかアルゴリズムを考慮する必要がある。

#### 参考文献

- [1] 富野剛, 井上亮文, 市村哲, 松下音: 多数人数参加型テレビ会議システムにおける発言者拡大映像の作成, 情報処理学会論文誌, vol.47, No7, pp.2091-2098 (2006).
- [2] 竹前嘉, 大塚和弘, 武川直樹: 対面の複数人対話を撮影対象とした対話参加者の視線に基づく映像切替方法とその効果, 情報処理学会論文誌, Vol.6, No.7, pp.1752-1767 (2005).