

高齢者の社会参加のために: 拡張リアルタイムコミュニケーションプラットフォームの提案

Facilitating Elderly Participation in Communities on a Standard-Based Extendable Platform for Augmenting Real-Time Communications

小杉晋央*¹
Akihiro Kosugi

西口昇吾*²
Shogo Nishiguchi

小林正朋*¹
Masatomo Kobayashi

*¹ 日本アイ・ビー・エム株式会社 東京基礎研究所
IBM Research - Tokyo

*² 大阪大学大学院基礎工学研究科
Osaka University Graduate School of Engineering Science

Achievement of real-time and rich remote communication is indispensable for people who face barriers to travelling due to aging. Technologies in this area are evolving toward not only imitating face-to-face communication but also augmenting users' cognitive experiences. However, most of them are built without a common interface or platform on which they can easily be integrated. We propose a standard-based platform for real-time communication in which features are componentized and bound with each other flexibly. This paper describes the mechanism of the platform and some applications as embodiments of the concept.

1. はじめに

自分が他者に必要とされていると感じることは、人が健康的に生きていく上で重要な要素である[1]。就労をはじめとした社会参加により、他者に何らかの価値を提供し、またそのことに意味を見出せるようになるには、相手との関係に実感を持つことが必要であり、このために人は、共同作業を始めるときに、顔合わせミーティングなどを行うことが多い。実際に顔をあわせた時の存在感により、相手を実在する人間関係の中で理解しやすくなるためである。高齢者など移動時の制約が伴う人々にとって、遠隔で行うしかなかったコミュニケーションにおいて自分と相手の存在感を伝えることは、大きな課題であったが、電話の発明以来、その手段は多様性を伴ってきている。ビデオ通話によるリアルタイムな映像の共有、テレプレゼンスロボットによる身体性の追加など、その技術は顔を合わせた状況を再現する形で発展してきたが、一方で、昨今では顔認識、表情モーフィング、センサー情報など、人以上に状況を把握、表現する技術も実現しつつある。

しかしながら、それらの技術は独立して開発、提供され、相互に利用可能な共通の枠組みができていない。このため、例えばテレプレゼンスロボットに対し、使いやすい操作の仕組みを開発しても、それを他のロボットに適用する場合に、まったく新しい枠組みで開発を行う必要がある。

そこで我々は、共通のユーザーシナリオである、リアルタイムなコミュニケーションを、遠隔で実現するための拡張可能かつ、なるべく広い範囲で利用可能なプラットフォームを提案する。本プラットフォームでは、Peer to Peer (P2P)通信による高速な接続を確保しつつ、追加機能がプラグインとして開発され、フレキシブルに再利用される。本稿では、そのプラットフォームの仕組みと、実装例について記述する。

2. WebRTC によるリアルタイムコミュニケーションプラットフォーム

本プラットフォームは、WebRTC[2]を使用した P2P 通信を確立し、複数の遠隔地間の映像、音声を含んだ様々なデータの高速なやり取りを管理する機能、やり取りしたデータに基づきブラウザ外のコンポーネントにアクセスする機能を持つ。プラットフォームは、それら個別の機能を Publish-Subscribe pattern[3]でインターフェースすることにより、アプリケーションが、機能の実装を緩やかに組み合わせることで実現することが可能となる(図 1)。Publish-Subscribe pattern は、jQuery[4]、Dojo Toolkit[5]など広く使われる JavaScript ライブラリに標準搭載のイベントシステムで実装可能である。これにより、映像や音声のみならず、ロボット操作、センサーの読み取りといった様々な機能、それに伴う多様なデータ形式を扱うコンポーネントに合わせた実装を、拡張しながら組み合わせることが可能となり、同時にそれらは P2P 通信による遠隔地でも遅延の少ない形での運用が可能となる。

実際に、本プラットフォームは、顔検出機能とロボットの遠隔操作機能を持つコンポーネントを組み合わせ、遠隔地間でののぞき窓操作モチーフを実現することによって、手を使わずに操作可能なテレプレゼンスロボットシステム[6]を作成する際に用いられ、半年に渡る高齢者同士の遠隔講習において運用された。

*¹ 小杉晋央, 日本アイ・ビー・エム株式会社 東京基礎研究所,
東京都中央区日本橋箱崎町 19-21, aikosugi@jp.ibm.com

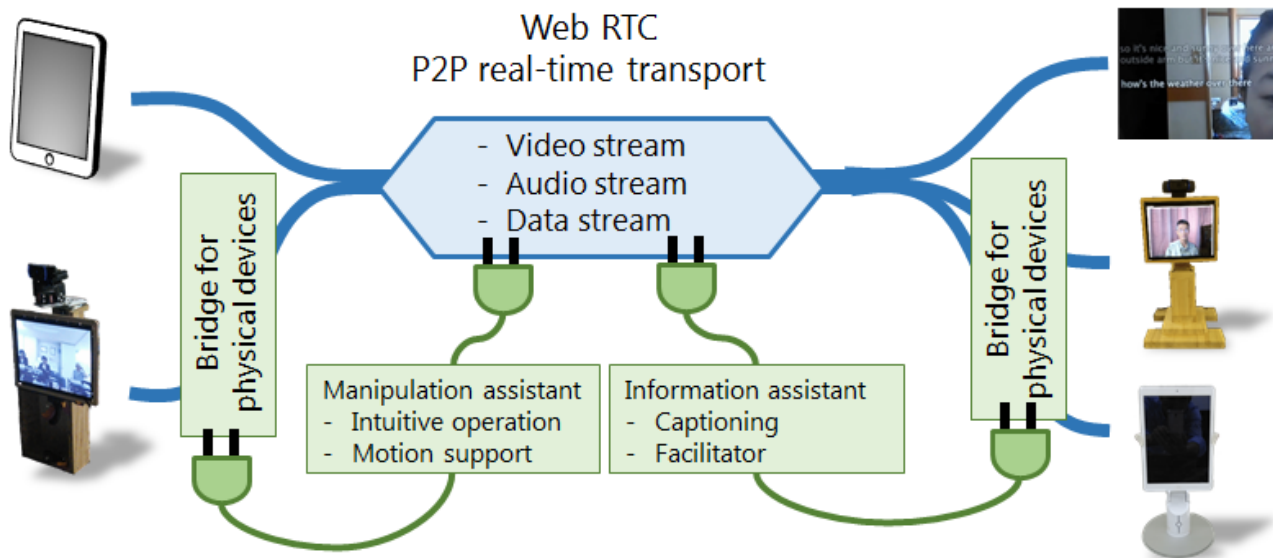


図 1 プラットフォーム全体図
Figure 1. Diagram of the platform

3. アプリケーション

提案プラットフォームを利用し、人同士のリアルタイムなコミュニケーションを補助、あるいは拡張するアプリケーションを例示する。

3.1 分散音声認識による会話内容の視覚化

遠隔会議における音声データをテキスト変換することにより、参加者個別の会話の内容を時系列のテキストデータで把握することが可能となる。この機能を利用して、高齢者を含め、音声聞き取りにくいユーザーのための、リアルタイムなキャプション、および、発話内容をタグクラウドとして表現することによる、思い出し支援機能を実装した(図 2)。

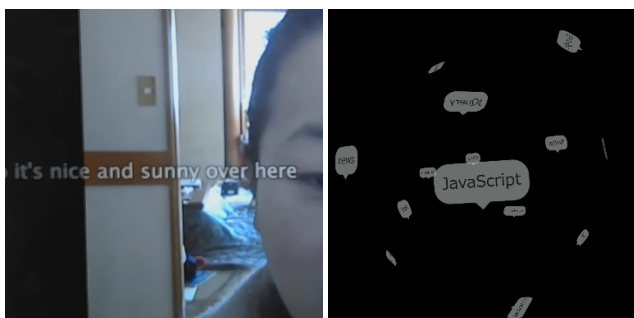


図 2 参加者発話のキャプション(左), 会話内容のタグクラウド(右)

Figure 2. Captioning participant's talk(Left), Tag cloud from conversation(Right)

これらの機能は、音声認識モジュールを共通コンポーネントとして通信モジュールから音声データチャンネルを Subscribe し、それぞれの視覚化モジュールが認識結果を Subscribe することにより、複数の機能間で実装と処理結果の共通化を可能とした(図 3)。

遠隔会議システムの一つのメリットは、発話者が個別の回線で参加するため、音声と発話者の同定が行いやすいことである。しかしながら一方で、音声認識エンジンが理解可能なデータ形式に変換する処理は、ある程度のマシンパワーを必要とし、ビデオ会議のように映像のやり取りも行いながら、複数の参加者の音声変換を同時に一つのマシンで行うことは、往々にしてパフォーマンス上の負荷を伴う。このため提案プラットフォームを使い、音声のテキスト変換の要求コマンド、および、その結果を P2P データチャンネルを用いてやりとりし、各ユーザーの接続したマシンで音声のデータ変換処理を代替で行い、処理を分散させることにより、複数参加者の会話内容の認識をリアルタイムに行った。このようなパフォーマンス上の問題は、今回の例を含め、開発がある程度進んだ段階で発見されることが多い。この機能を作成した際も、当初は視覚化を行うマシンで全ての処理を行う設計で実装を始めたが、会議参加者数が増えるにつれ、音声データ変換の遅延による品質の低下が発生したため、当プラットフォームの通信コンポーネントに音声認識コンポーネントのコマンド体系モジュールをプラグインすることにより、実装の枠組みを大きく変えることなく、問題への対処が可能となった。

3.2 遠隔講義支援「バーチャルエージェント」Mocoro

テレプレゼンスロボットなどの導入により、遠隔の講習でも講師の存在感を伝えられるようになりつつあるが、生徒の様子、教室の雰囲気を講師に伝えることは難しい。また、高齢者など、スキルを持つ人が、教えることに必ずしも長けているとは限らない。そこで、遠隔でのやり取りをリアルタイムに把握できる本プラットフォームを利用して、生徒と講師の会話内容を把握し、生徒の雰囲気を自然な形で講師に伝えたり、生徒の理解度を考慮した会話の仕方を自然に講師に誘導するための自律ロボット、Mocoroを開発した。(図 4)

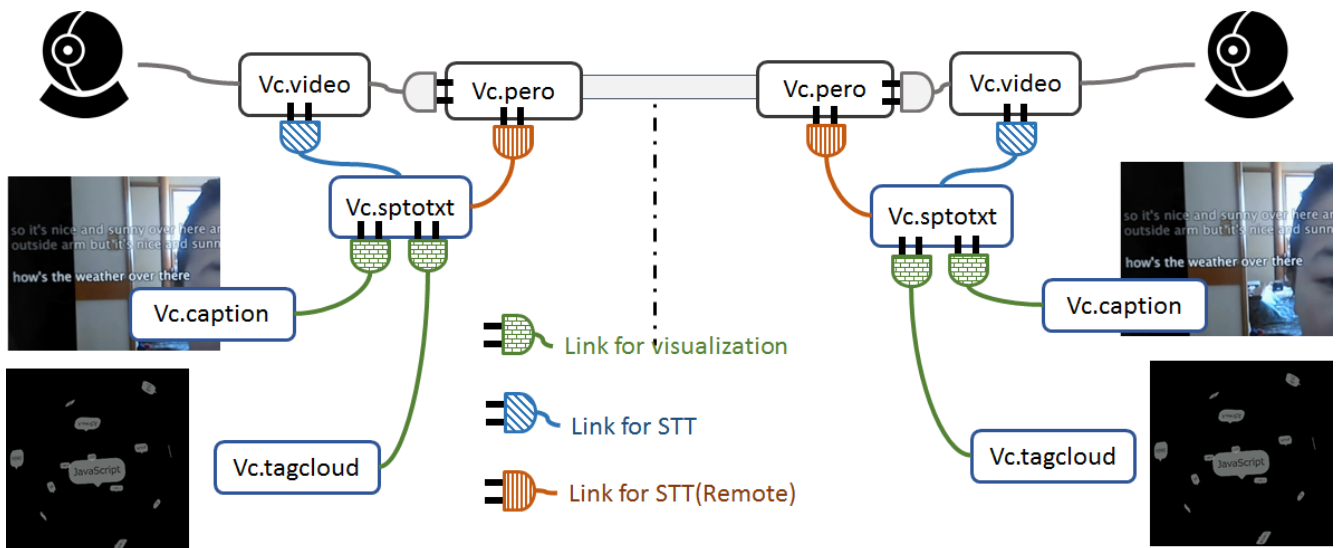


図3 会話内容視覚化におけるコンポーネントの関係

Figure 3. Diagram of the components for conversation visualization

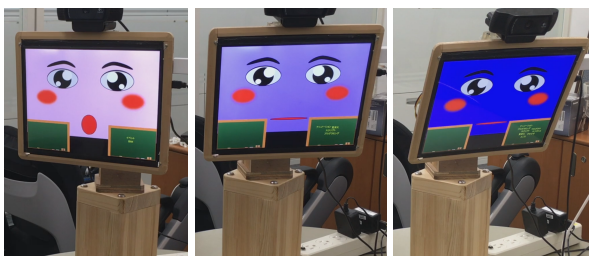


図4 バーチャルエージェント“Mocoro”

Figure 4. Virtual Agent “Mocoro”

Mocoro は、pan-tilt 機能を持った卓上テレプレゼンスロボットに搭載され、本プラットフォームでの会話内容の視覚化で追加された音声のテキスト変換モジュールにより、参加者の話を聞けるほか、テキスト読み上げサービスを利用して、あらかじめ用意した組み合わせによる受け答えが可能である。さらに、講師の発話スピードや言い淀みの頻度によって、表情や振る舞いを変化させ、講師が会話の進め方について気づき、修正することをサポートする。例えば、秒間にテキスト変換された文字数を話速とし、NHK アナウンサーの会話スピードである秒間 5 文字を基準に、早すぎる場合や遅すぎる場合、あるいは、“えーと”、“あー”、“そのー”といった言い淀みの時間あたりの回数をカウントし、その数が多い場合に、徐々に俯き加減になり、顔色を青くすることで、聞き手が混乱している表現を行う。ロボットは、講師あるいは、生徒たちの側に置かれ、孤独になりがちな遠隔授業のアシスタント、あるいは、教室での一生徒のような形で遠隔授業に参加する。警告を表現するにあたり、実体を伴いながらもそばで動作表現を行うことは、音や光、あるいは、操作画面上の画像を用いることに比べて、授業のやり取りに対するインタラプトの少ない緩やかなやり方であり、また、講師が遠隔ではなく、face to face で授業をしている環境と同じように、自然に生徒の様子として気づくことが可能と期待できる。

4. まとめ、展望

本稿では、リアルタイムな遠隔コミュニケーションのための、拡張可能なプラットフォームと、その応用の可能性について述べた。

PC や、スマートフォンなど、人が持つデバイスの処理能力やネットワークの回線幅が向上してきた昨今、リアルタイムな遠隔コミュニケーションにおいて、多様な情報をやり取りすることが現実のものとなってきている。ロボットなどの物理デバイスを用いた実体化など、情報の伝達及び表現手段も多岐にわたって利用可能になってきた。このような状況下で、リアルタイムコミュニケーションのためのアプリケーションの開発、および新たな技術の導入コストの低減を意図した提案プラットフォームの果たす役割は大きいと考える。今後、現実のコミュニケーションをさらに補う様々な機能が実装され、人々の意思疎通手段が距離を超えたリアルタイムコミュニケーションにシフトしていくことで、高齢者を含めた、これまで移動時の制約により社会参加を阻害されていた人々が、分け隔てなく社会参加することができ、いつでも自立した立場で生活できるようになることを望む。

参考文献

- [1] Liu: “Helping Others Make Me Happy”: Social Interaction and Integration of People with Disabilities., CSCW'16, ACM, 2016
- [2] Web-RTC: <https://www.w3.org/TR/webrtc/>
- [3] Publish-subscribe pattern: https://en.wikipedia.org/wiki/Publish%E2%80%93subscribe_pattern
- [4] jQuery: <https://jquery.com/>
- [5] Dojo Toolkit: <https://dojotoolkit.org/>
- [6] Kosugi: Hands-Free Collaboration Using Telepresence Robots for All Ages., C SCW'16, ACM, 2016