

ConvivialChat: テキスト参加者とスピーチ参加者の対等で 相互補完的なコミュニケーションを可能にするシステムの提案

藤井孝弘¹ 西田健志²

概要: 現在、主流として使用されている遠隔会話ツールは、音声による通話機能の他にテキストによるチャット機能も備えており、主に音声通話を利用する参加者とチャットを利用する参加者が混在した状況を可能にしている。しかし、既存の遠隔会話ツールを用いて行われる話し合いの場にチャットのみで参加した場合、音声による参加者同士でリアルタイムに進行する会話へ積極的に貢献することは難しく、疎外感や発言への抵抗を覚える可能性がある。そこで我々は、チャットに送信したテキストが機械音声で読み上げられる機能、気軽な反応を可能にする音声付きのリアクションボタン、音声発言がチャットのログに表示される機能といった、チャットによる参加者と音声通話による参加者が互いに対等で相互補完的な話し合いを行うための諸機能を備えたシステム ConvivialChat を提案する。実験の結果、チャットによる参加者が積極的に話し合いに貢献することができる可能性が示されたが、当初の目的に逆行する効果を示した機能もあり、各機能の目的の再検討やユーザインタフェースの改善などの課題が明らかになった。

キーワード: チャット、音声通話、音声認識、音声合成、テキストとスピーチの共生

ConvivialChat: A Remote Conversation Tool Designed for Text-Speech Symbiosis

TAKAHIRO FUJII^{†1} TAKESHI NISHIDA^{†2}

Abstract: Current mainstream tools for remote conversations, not only have functions to allow communication through voice, but also have text-chat functions. This allows the mixing of participants who speak as the main way of communicating and those who prefer to use only text. However, in the case where an individual decides to communicate only through text using pre-existing tools for remote conversations, it is difficult to actively contribute in discussions with others that communicate verbally. There is also the possibility of the non-verbal participants experiencing feelings of alienation and hesitation to make their thoughts known. We propose a new system named ConvivialChat which would provide a way for participants to have equal and complementary discussions with each other. This is made possible through built-in functions such as the reading of text by a synthesized voice, voiced reaction button, and a function to turn voiced remarks into text which would be written into the chat log. Experiments conducted showed the potential for individuals who exclusively use chat to readily join in conversations with ConvivialChat. The results also showed the effects for one function, in particular the feature that turns speech into text, did not align with the original purpose of its design. Thus, illustrating the need to reexamine the purpose of each function and to improve the user interface.

Keywords: chat, tele-communication, text, speech, speech-to text text-to-speech text -speech-symbiosis

1. はじめに

現在、主流として使用されている遠隔会話ツールは、音声による通話機能の他に、テキストによるチャット機能も備えており、主に音声通話を利用する参加者（以下スピーチ参加者と呼ぶ）と主にチャットを利用する参加者（以下テキスト参加者と呼ぶ）が混在した状況を可能にしている。例えば、電車やバスなどの公共交通機関で移動中である場合、スピーチ参加者として話し合いに参加することは困難であるし、その他にも周囲の騒音が気になる環境から参加する場合は、マイクをオンにすることが憚られ、テキスト参加者としてチャットのみを使用することが十分に考えられる。COVID-19の影響で、遠隔会話ツールの使用機会が増加・多様化したことで、テキスト参加者とスピーチ参加

者の混在した話し合いの場面は増えていると考えられる。

しかし、既存の遠隔会話ツールにおいてテキスト参加者とスピーチ参加者により行われる話し合いの場には、互いの円滑なコミュニケーションを妨げる問題が存在する。まず、テキスト参加者がチャットに送信したメッセージが、リアルタイムに進行するスピーチ参加者同士の会話に取り上げられにくいことが挙げられる。テキスト参加者の発言が、話題に上るかどうかはスピーチ参加者次第であり、テキスト参加者はスピーチ参加者と対等に話し合いに貢献することができない状況があると考えられる。また、スピーチ参加者同士は、相槌のような声による気軽なリアクションを行うことができるのに対して、テキスト参加者は気軽な反応をスピーチ参加者と同じように示すことができない。テキスト参加者のリアクション方法として、既存の遠隔会

¹ 神戸大学 国際人間科学部
Kobe University, Department of Global Human Sciences

² 神戸大学大学院国際文化学研究科
Kobe University, Graduate School of Intercultural Studies

話ツールにおいてもスタンプを表示する機能が用意されているが、これらの機能による意思表示には視認性の低さが指摘されている[1]。さらに、話し合いの場において少数派であると考えられるテキスト参加者は、自分のメッセージやリアクションが気付かれているかどうか分からない状況に疎外感を感じてしまい、チャットにメッセージを送信すること自体に抵抗を覚えてしまうことも考えられる。多様性の重要性が叫ばれる現代であるにも関わらず、増加傾向にあるオンラインでの話し合いの場面において、テキスト参加者とスピーチ参加者の間に分断が生じてしまうような事態は望ましくない。

そこで本研究では、テキスト参加者とスピーチ参加者の混在した状況を想定し、互いに対等に話し合いに参加できる分断のない相互補完的なコミュニケーションを実現するためのシステム ConvivialChat を提案する。このシステムの主な機能は3つだ。1つ目は Text to Speech 技術によりチャットに送信したテキストが機械音声で読み上げられる「TTS 機能」。2つ目は、顔文字のスタンプが送信されると同時にその表情に即した音声を読み上げられる「リアクションボタン読み上げ機能」。そして、3つ目は Speech to Text 技術により音声による発言がテキストとしてチャットのログに表示される「STT 機能」である。

これらの機能を備えたシステムの効果を検証するために実験を行った。実験は、提案システムの主な3つの機能の有効性を検討するために Baseline システムとの比較実験を行い、実験後にアンケートを実施した。実験の結果、TTS 機能とリアクションボタン読み上げ機能は当初の期待通りの効果を確認できたが、STT 機能については期待した効果と正反対の効果を生む可能性が示唆された。また、それぞれの機能についての課題や目的の再設定の必要性、ユーザインタフェースの改善点などが明らかになった。

2. 関連研究

本章では、Text to Speech 技術を用いたコミュニケーション支援システム、Speech to Text 技術を用いたコミュニケーション支援の研究、テキストとスピーチの共生を志向した研究を紹介する。

2.1 Text to Speech 技術を用いたコミュニケーション支援

Commentable [2]は、遠隔授業を受ける学生が音による積極的なリアクションを可能にしたシステムで、テキストで送信されたコメントやリアクションボタンが送信と同時に機械音声によって読み上げられる。このシステムを用いてオンラインで発表会を行ったところ、音声ありのリアクションの方が多かったことが記されている。このシステムは、参加者同士の音声会話を中心としたリアルタイムの話し合いの場を想定する本研究と使用場面は異なるが、TTS 機能とリアクションボタン読み上げ機能という本研究の提案システムに実装した機能と同一の機能を備えている。

2.2 Speech to Text 技術を用いたコミュニケーション支援

Inspiration Wall [3]は、認識した音声からキーワードを抽出し、ディスプレイ上にそれらをワードクラウドとして表示することで、2人1組の対面のアイデア出し会議の支援を行った研究である。実験の結果、実験参加者がディスプレイを見る回数が多いグループほど多くのアイデアを出すことがわかった。本研究では、スピーチ参加者とテキスト参加者を対等に扱う手段として音声認識を用いているが、同様にアイデア出し等の刺激になる可能性は考えられる。

See-Through Captions [4] は認識した音声を透明のディスプレイにそのまま表示することで、聴覚障害者と健常者のリアルタイムのコミュニケーションを支援する研究である。本研究では障害者支援を主目的とはしていないが、テキストによるコミュニケーション参加を円滑にすることは障害者支援にもつながる可能性が高いと考えている。

2.3 テキストとスピーチの共生を志向した研究

EnhancedChat [5] は、学会やシンポジウムのような場面でチャットと音声会話の関連性を保つことができていることに課題意識を持ち、実空間の音声による会話とチャットのやりとりを関連付けることのできるシステムを提案した研究である。

ChaTEL [6] は、音声による会話とチャットのコミュニケーションの差違に注目し、音声を用いてリアルタイム性を保持しながらマルチスレッド対話を実現するために、録音した音声データを送り合うことができるチャットシステムを提案した研究である。さらに、ChaTELex [7] の研究では、ChaTEL に通常のテキストによるチャット機能を加え、「音声とテキストが混在する」システムが検討されている。

本研究では、TTS 機能や STT 機能によりテキストとスピーチを共通のチャットのログに共生させ、1章で述べたテキスト参加者の課題を解決することで、テキスト参加者とスピーチ参加者という2種類の参加形態による対等で相互補完的なコミュニケーションの実現を目指す。

3. ConvivialChat のデザイン

本章ではまず、システムのデザイン指針を記す。その後、提案システムのユーザインタフェースとデザイン指針に沿って実装した特徴的な3つの機能を説明する。

3.1 デザイン指針

既存の遠隔会話ツールに存在する問題を踏まえ、テキスト参加者とスピーチ参加者の対等な話し合いへの参加を実現するために、3つのデザイン指針を定めた。

- (1) テキスト参加者とスピーチ参加者の発言を平等に周知する。
- (2) テキストとスピーチの双方による発言が対等に確認できる機能とインタフェースを目指す。
- (3) テキスト参加者の参加形態に起因する疎外感や発言への抵抗を軽減し、気軽な話し合いへの参加を促す。

3.2 ユーザインタフェースと機能

ConvivialChat は、テキストによるチャット上の会話と音声による会話の双方が行えるよう音声通話とチャットの機能を備えたウェブアプリケーションである (図 1)。

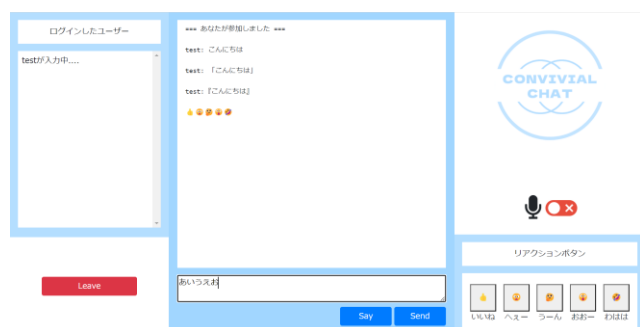


図 1 ConvivialChat のユーザインタフェース

左上にはユーザーのログイン名が表示される。チャットを送信しようとテキストボックスに文字を打ち込んでいるユーザーの名前の横には「(ログイン名) が入力中...」と表示される。テキスト参加者が話し合いから取り残されているという疎外感を抱くことがないように自分の発言の意思を気付いてもらいやすくするために実装した。

チャットの送信方法には、Send ボタンと Say ボタン (後述) の 2 種類があり、Send ボタンを押すと、通常のチャット送信が行え、テキストボックスに打ち込んだ内容がログイン名と共にチャットのログに表示される。

右下には、テキスト参加者が単純な反応を行うための手段としてリアクションボタンを実装した。リアクションボタンの種類としては、「いいね」「へえー」「うーん」「おおー」「わはは」の 5 つが用意されている。

遠隔会話ツールにおいてテキストとスピーチの双方による対等で相互補完的なコミュニケーションの場を探索するという目的に焦点をあてて実験・考察を行うために、チャットをメインとした遠隔会話ツールにおけるビデオウィンドウへの注目度の低さ[8]も踏まえ、ビデオ機能は実装しなかった。

3.2.1 TTS 機能 (Say ボタン)

TTS 機能は、テキスト参加者がチャットで送信したテキストが機械音声で読み上げられる機能である。既存の遠隔会話ツールでは、チャットのメッセージが会話に拾われるかどうかはスピーチ参加者次第であるし、テキスト参加者は自分の発言が全員に視認されたうえでスルーされたのか、それとも見られていないのかを判断する術はなく、スピーチ参加者と対等な立場で話し合いへ参加することができるとは言い難い。音声の読み上げを行えば、テキスト参加者の意見も確実にスピーチ参加者に届き、テキスト参加者とスピーチ参加者の発言を平等に周知できると考える (デザイン指針 1)。

3.2.2 リアクションボタン読み上げ機能

リアクションボタン読み上げ機能は、顔文字のスタンプの送信と共に、その表情に合った音声を読み上げられる機能であり、主にテキスト参加者による使用を想定している。スピーチ参加者の場合、相槌により会話を進めようという意思を簡単に示し、会話を円滑に進めることができるが、テキスト参加者が同様の気軽な反応を示すことは難しい。既存の遠隔会話ツールにはテキスト参加者のリアクション手段としてスタンプ表示機能が備わっているが、音声中心の会話が進行する状況で、スタンプという視覚に訴えるリアクション自体が最適なアプローチであるとは考えにくい。この機能を用いることで、音声を伴うリアクションによるテキスト参加者の気軽かつ積極的な会話への貢献が可能になると考える (デザイン指針 3)。

3.2.3 STT 機能

STT 機能は、スピーチ参加者の発言を音声認識機能によってテキスト参加者と共通のチャットのログに表示する機能である。STT 機能のオンオフと、マイクのオンオフは連動している。スピーチ参加者の発言がチャットのログに表示されれば、自ずとチャットのログに意識が向き、チャットの発言を視認する機会が増加すると考えられる。また、テキスト参加者とスピーチ参加者の双方の発言がすべて共通のチャットのログに表示されることで、話し合いへの参加具合を互いに対等に確認しあうことができ (デザイン指針 2)、テキスト参加者の発言だけがチャットのログに表示されることに対する疎外感やメッセージ送信への抵抗を軽減することができるのではないかと考える (デザイン指針 3)。これらの効果により、スピーチ参加者が主導権を握りがちなオンラインでのリアルタイムの話し合いにおいてテキスト参加者とスピーチ参加者が対等に話し合いを進めることができると考える。加えて、過去のログを参照した話題転換の可能性や、事後の振り返りの際の材料としての議事録の効果も期待できる。

3.3 想定する使用方法

ConvivialChat の使用場面については、企業のチームミーティングやサークルの運営ミーティングなどオンラインで参加者同士が対等な立場で話し合いを行う場面が考えられる。また、音声発話を中心に行う参加者 (スピーチ参加者) と、社会的/身体的その他さまざまな都合により音声発話を行うことが難しい参加者 (テキスト参加者) の双方が混在しており、音声発話を中心に会話が進行する場面を想定している。

スピーチ参加者は、会話ページに入った後マイクをオンにし、主に音声によるコミュニケーションを行う。遠隔会話ツールで話し合いを行う場合に最も想定されるだろう参加形態である。2 種類のチャット送信とリアクションボタンの使用については特に禁止されているわけではない。

テキスト参加者は、会話ページに入った後、マイクをオフにしたままの状態で話し合いに参加する。2種類のチャット送信とリアクションボタンを適宜使用して、音声コミュニケーションに積極的に参入していくことが想定される。双方の参加形態について表1に整理した。

表1 参加形態と使用機能の整理図

	チャット送信	マイク	リアクションボタン
テキスト参加者	◎	×	◎
スピーチ参加者	○	◎	○

◎→主な使用を想定 ○→特に使用を禁止されない
×→使用不可

4. 実験

TTS機能、リアクションボタン読み上げ機能、STT機能という特徴的な3つの機能を備えたConvivialChatを使用することで、テキスト参加者とスピーチ参加者の双方が話し合いを行う場において、実際にはどのようなコミュニケーションが展開されるのか・実装した機能には期待した通りの効果があったかを検証するために実験を行った。

4.1 実験の概要

本実験は、あるお題について5人組で話し合いをするオンライン実験である。ConvivialChatの機能の効果の検証のため、ConvivialChatに実装した3つの機能（TTS機能、リアクションボタン読み上げ機能、STT機能）を備えていないBaselineシステムを用意して比較することとした。表2にConvivialChatとBaselineシステムの違いを示す。Baselineシステムにおいて、チャットの送信の手段はSendボタンのみであり、リアクションボタンはスタンプがチャットのログに表示されるだけで機械音声による読み上げは行われず、STT機能によって音声発言がチャットのログに表示されることなく音声通話を行うことのできるシステムである。

表2 ConvivialChatとBaselineシステムの比較

	チャット送信ボタン	リアクションボタン	STT機能
ConvivialChat	SendとSay	読み上げあり	あり
Baseline	Send	読み上げなし	なし

両システムの比較の検証方法について、まずは実際にどのようなコミュニケーションが展開され、システム間でどのような違いがあったのかを観察・考察するために、実験中の画面を録画し、音声を録音した。また、各機能に期待した効果が得られた場合に変わると考えられる項目やシス

テムの使いやすさ、おもしろさなどを尋ねる7件法のアンケートを行った。さらに、それぞれの機能についての感想や意見を自由に尋ねるアンケートを実施した。

4.2 実験参加者

実験には、計15人（女性4人、男性11人、平均年齢21歳、母語が日本語：14人、日本語と英語のバイリンガル1人）が参加した。参加者は筆者と同じ学部所属している学生であった。実験参加者を募るために、学部2年生向けの講義である「プログラミング基礎演習」で、実験参加の呼びかけを行った。また、著者の友人で同講義の受講経験のある者に実験参加を依頼した。さらに、新入生オリエンテーションという著者が所属する学科の公式行事で企画を共同で行ったメンバーのLINEグループで実験参加を呼びかけた。

参加者を5人ずつの3グループに分け、3回の実験を行った。3グループ中2グループはプログラミング基礎演習の受講者および受講経験のある者（学部4年生4人、学部3年生2人、学部2年生4人）で構成されており、もう1グループは新入生オリエンテーションを共同で行ったメンバー同士（全員が学部4年生）で構成されていた。

4.3 実験の手順

実験はオンラインで行われ、実験参加者は各自任意の場所から各々のパソコンで参加した。実験の手順の概要を図2に示す。実験の前半は、Baselineシステムを使用した15分間のお題に沿った話し合い後に7件法のアンケート、後半はConvivialChatを使用した15分間のお題に沿った話し合い後に7件法のアンケートとConvivialChatを対象にした自由記述ベースのアンケートを実施した。

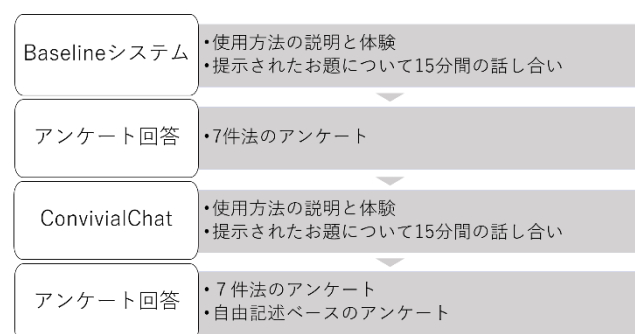


図2 実験の手順の概要

表3に各グループが扱った話し合いのお題を示す。お題は3つ用意し、各グループで異なる2つのお題について話し合ってもらった。カウンターバランスを考慮し、それぞれのお題が異なる条件下で扱われるようにした。用意したお題は、学科公式行事という共通の話題、受講経験のある共通の授業、コロナ禍という状況での共通の話題としての話しやすさを考慮し、「これからの新入生オリエンテーションはどうあるべきか」（新オリ）、「授業後もプログラミング

の勉強を続けるにはどうすればいいか」(プログラミング),
「オンラインで新しい友達をつくるにはどうすればいいか」
(新しい友達) の3つである。

表 3 グループ毎に扱ったお題

	Baseline システム	ConvivialChat
グループ 1	プログラミング	新しい友達
グループ 2	新オリ	プログラミング
グループ 3	新しい友達	新オリ

4.4 結果

まず、実験の様子を録画した動画を観察した結果について、チャットのメッセージが会話に反映されていたか、リアクションボタンの会話への影響、機能に対する言及の3つの観点から記述する。その後、実験後に行った2種類のアンケート結果を記す。

4.4.1 観察結果

(1) チャットのメッセージが会話に反映されていたか

Baseline システム使用時には、テキスト参加者のメッセージがスピーチ参加者に取り上げられる場合、取り上げられない場合の両方が観察された。

取り上げられる場合については、テキスト参加者の発言に対してすぐにスピーチ参加者が相槌を打つ様子が見られた。その他、発言中だったスピーチ参加者がチャットのメッセージが更新されたことにすぐに気づき、話すのを止めて反応する場面や、発言中の参加者とは別の他のスピーチ参加者がすぐにチャットのメッセージに気づき反応する場面があった。さらに、テキスト参加者が送信したメッセージが時間差でスピーチ参加者によって言及され、会話が進んでいる様子も見られた (図 3 上)。

一方で、スピーチ参加者同士の会話が絶え間なく行われ話が進んでいる最中にテキスト参加者が送信したメッセージが、スピーチ参加者の会話に取り上げられることなく話し合いが進行する様子が見られた。

CovivialChat 使用時にも、Send ボタンで送信されたメッセージがスピーチ参加者に取り上げられる場面、取り上げられない場面があった。

Baseline システム同様、テキスト参加者の発言に対してスピーチ参加者がすぐに反応を返す様子がみられたものの、スピーチ参加者同士の話し合いが絶え間なく継続しているときに送信されたメッセージは会話に取り上げられずに流される場面もあった。

S1: 『新オリでその後よく話す友達にあえたとかありましたか？私は特になくて』

S2: 『私もなかったです』

S3: 『新オリ開催後1ヶ月は、割と新オリでできた友達と交流するってことは合ったと思うんですけど (以下略)』

S4: 『なるほど』

S1: 『S4ちゃんはどうやった？』

T: よっともが増えた感じ！

S4: 『私は、オンラインでやったやつは～友達はできてなくて (以下略)』
(Tの発言には触れず、S1の質問に答える)

～約1分40秒間スピーチ参加者同士で話が続く～

S1: 『Tさん、よっともが増えた感じって』
(テキスト参加者の発言に言及)

～ここから約1分間、よっともについて話が広がる～

S2,S4: ～ (二人でSTT機能の誤字認識を指摘し合う)

S2: 『全部変換されるの怖いですね』

T: 「なんか集まるなら共同作業したいですね」
(テキスト参加者がお題に沿った発言)

S2: 『そうですよねやっぱオンラインでさ友達になろうって言っても厳しいものがありますね』
(Sayボタンの発言に反応し、話し合いが再開)

図 3 実験における特徴的な会話場面の抜粋

上: Baseline 条件 (お題「新オリ」)

下: ConvivialChat 条件 (お題「新しい友達」)

S1~4: スピーチ参加者 T: テキスト参加者

『』: 音声発言 「」: Say ボタンによるテキスト送信

平文: Send ボタンによるテキスト送信

一方で、Say ボタンで送信されたメッセージに対しては、どのグループの話し合いにおいても基本的にすぐにスピーチ参加者から反応が返され、そのメッセージを受けて話し合いが進行する様子が見られた (図 3 下)。ただ、スピーチ参加者の発言中に Say ボタンでメッセージが送信され読み上げ行われた際に、その発言をスルーして話し合いを進める様子も1度だけ見られた。しかし、同グループのその後の話し合いで Say ボタンによる質問の体を表した発言が、スピーチ参加者の発言と重なった際には、スピーチ参加者が発言を継続しながらもその最中に、質問への回答と思われる発言が行われ、テキスト参加者からのメッセージが話し合いに反映されていた。

(2) リアクションボタンの会話への影響

Baseline システムについては、リアクションボタンの使用がたびたび見られ、スタンプが連続で送られているのに対してスピーチ参加者が反応する様子が2度ほど見られた。

ConvivialChat についても、リアクションボタンの使用が何度も見られた。機械音声の読み上げが行われると、それに対して最初は何のグループも苦笑が起き、会話がそこで途切れてしまう様子だった。ただ、話し合いが進行するにつれて、読み上げ機能をスルーして会話が進む様子も見られた。スピーチ参加者が会話中に「わはは」のスタンプが送信され、それに対して返答する様子も見られた。

加えて、3 回目の ConvivialChat の実験の最後には、リアクションボタンを複数人で連打しあっておもしろがる様子が見て取れ、その際、読み上げ処理が溜まってしまい、その後の Say ボタンの発言が読み上げられるまで時間がかかってしまう場面もあった。

(3) 機能に対する言及

ConvivialChat 条件では、テキスト参加者が、「送るタイミングミスりました笑」や「なんか Say 使うの抵抗ある笑」というメッセージを送信し、TTS 機能への違和感を示す様子があった。また、スピーチ参加者が STT 機能に対して「全部変換されるの怖いですね」(図 3 下)や「これ失言できないですね」など同じく機能に対する違和感を示す場面が見られた。また、音声認識誤りに対して、指摘しあったり笑い合ったりする様子がどのグループにも見られた。

4.4.2 アンケート結果

表 4 に 7 件法のアンケートの内容と結果の詳細を示す。参加者 15 名のアンケート結果をサンプルデータとして、Baseline システムと ConvivialChat のアンケート結果に対応のある両側 t 検定を行ったところ、「チャットの履歴は見やすかったですか?」と「このシステムを使ってみて、使いやすいと思えましたか?」の 2 個の項目に有意差が見られた ($p < 0.05$) ($p: 0.00, 0.00$)。また、有意差を得られた 2 つのアンケート項目の相関係数は 0.4 であった。

3 つの機能の必要性について、「はい/いいえ/わからない」の 3 択で尋ねた結果を図 4 に示す。以下、それぞれの回答の理由を自由記述で尋ねた質問への回答を各機能別に記す。

(1) TTS 機能 (Say ボタン)

TTS 機能は不必要だと答えた側の意見として「読んだ方が早い。会話の中に急に機械の声が入ってくるとびっくりする。」という回答が得られた。

必要だと答えた側の意見としては「緊急でチャットを見て欲しい時には使えるのでは普段は必要ないと思うが、重要なチャットを全然みない人がある時に、意思表示として使えると思う」やテキスト参加者からの「会話の主導権を握れるから」といった回答を得た。

表 4 7 件法のアンケートの内容と結果

(B : Baseline システム, CC : ConvivialChat)

質問項目	B 平均値 (標準偏差)	CC 平均値 (標準偏差)
発言はしやすかったですか? (テキスト参加者はチャットでの送信がしやすかったですか?) (7: とても気軽にできた)	5.6 (1.4)	5.0 (1.6)
発言するタイミングは取りやすかったですか? (テキスト参加者はチャットで送信するタイミングが取りやすかったですか?) (7: とても取りやすかった)	4.5 (1.9)	3.9 (1.8)
チャットの履歴は見やすかったですか? (7: とても見やすい)	6.3 (0.7)	3.1 (1.6)
チャットの履歴はどれくらいの頻度で見ましたか? (7: ほぼ全部見た)	6.1 (1.3)	5.0 (1.6)
このシステムを使ってみて、使いやすいと思えましたか? (7: とても使いやすい)	5.7 (1.2)	3.5 (1.5)
このシステムを使ってみて、おもしろいと思えましたか? (7: とてもおもしろい)	5.9 (1.1)	6.2 (1.1)
自分の発言(テキスト参加者はチャットに送信したテキスト)にリアクションがあると感じましたか? (7: とても感じた)	5.5 (1.6)	5.7 (0.9)
話し合いに自分は貢献できたと思えますか? (7: とても貢献した)	4.9 (1.6)	4.9 (1.2)
会話から取り残されていると感じる場面はありましたか? (7: とても多くあった)	2.6 (1.7)	2.3 (1.1)

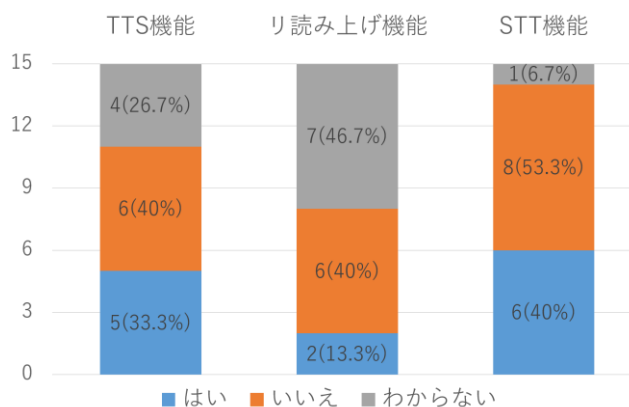


図 4 各機能の必要性についてのアンケート結果

わからないと答えた側の意見としては「TTS 機能による音声の流れるときにそちらに注意が向いてしまい、会話が一時的にストップしてしまうように感じた。しかし、会話に集中してテキストを読む暇がないときには、テキストの内容が読み上げられた方がいいなと感じた」というような利点と欠点の双方を挙げているものが見られた。

(2) リアクションボタン読み上げ機能

リアクションボタン読み上げ機能は不必要だと答えた側の意見として「～会話の妨げになったことがあった」や「見たらわかる絵文字なので、読み上げる必要がないと感じたから」などが見られた。

必要だと回答した側の意見としては「一瞬だけいいね！と言われて嬉しかったです」や「単純に顔文字だけだとこぼれ落ちる情報があるかもしれないと感じたから」などがあった。

わからないと答えた側の意見としては「～相槌代わりになるという面があった。～一方で、発言を遮ったり連打するとうるさくなるというデメリットがあり、気軽に押しづらい気はした」という意見が得られた。

リアクションボタンに関しては、ConvivialChat を用いた実験の際に使用したリアクションボタンの種類も尋ねた。その結果を図 5 に示す。

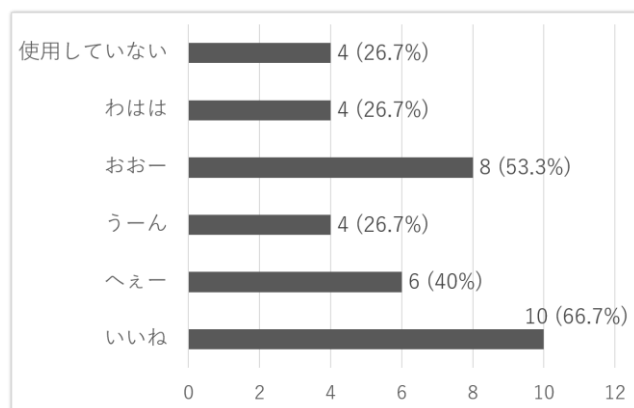


図 5 実験中に使用したリアクションボタンの種類

(3) STT 機能

STT 機能は不必要だと答えた側の意見として「自分の発言が文字化されることで発言内容に自信がなくなる、文字起こしの誤字が気になって集中できない、非発話参加者の発言が埋もれてしまうから」や「チャットのみで参加した人のチャットが流れてしまうから」などの回答が得られた。誤認識に対する否定的意見が多く見られた。

必要だと答えた側の意見としては「話したことが詳細に記録されていれば、後から会話の内容を見直したときに細かいところまでわかるところが便利だと感じた」や「zoom でよくある、発言が重なる場面で文字にされているとわかりやすいから」という回答を得た。

わからないと答えた側の意見として「面白い変換が会話の話題になるが、チャット本来の過去の発言を顧みるという点では、機能を果たしているとは言えないように感じたから」という意見もあった。

5. 考察

実験結果を踏まえ ConvivialChat に実装した特徴的な 3 つの機能に期待した効果の検証と、7 件法のアンケートで有意差のあった項目についての考察を行う。

5.1 各機能に期待した効果の検証

(1) TTS 機能 (Say ボタン)

TTS 機能によって期待される効果は、チャットのメッセージがスピーチ参加者同士の話し合いから置き去りにされることを防ぎ、テキスト参加者がスピーチ参加者と対等に話し合いに参加することを可能にすることであった。

観察結果や 2 つ目のアンケート結果からテキスト参加者の積極的な話し合いへの関与の可能性が示唆されたことから、この機能の当初期待していた効果がある程度得られたと考えられる。ただ、機械音声の不自然さの指摘も見られたことから改善の余地は残っていると考えられる。

(2) リアクションボタン読み上げ機能

リアクションボタン読み上げ機能によって期待される効果は、スピーチ参加者が行う相槌のようなリアクションをテキスト参加者が行うこと可能にし、テキスト参加者の積極的な話し合いへの参加と円滑なコミュニケーションを実現することであった。

2 つ目のアンケート結果からリアクションボタンの相槌の効果を示唆する意見も見られたことから、この機能の当初期待していた効果がある程度得られたと考えられる。

ただ、TTS 機能と同様、機械音声に改善の余地が残っているほか、リアクションボタンの種類に関して、本研究で用意したリアクションボタンは全て実験中に使用されていた (図 5) ことから、会話への貢献がある程度見込めるリアクションボタンを用意できたのではないかと考えられるが、学術的知見と今回の実験におけるリアクションボタンの使用のされ方も参考に、より最適なリアクションボタンの種類を検討する余地は残っていると考えられる。

(3) STT 機能

STT 機能によって当初期待される効果は、スピーチ参加者のチャットの視認機会の向上、共通のチャットのログを用いた話し合いへの参加具合の対等な確認、テキスト参加者の発言だけがチャットのログに表示されることに対する疎外感やメッセージ送信への抵抗感の軽減、過去のログからの話題転換、そして議事録としての効果であった。

アンケート結果や観察結果から、議事録的な効果の可能性が示唆されたものの、それ以外の効果については確認することができなかった。特に、チャットの視認機会の向上について、1 つ目のアンケート項目「チャットの履歴はどれくらいの頻度で見ましたか？」では有意差を得ることはできておらず、期待していた効果を得ることはできなかったと考えられる。また、テキスト参加者の疎外感やメッセージ送信への抵抗感を軽減することについては、テキスト

参加者のみのデータ数が少なすぎるため信憑性のある分析を行うことはできておらず、アンケート結果や実験の観察結果からもそのような効果があると判断できる意見やデータを得ることはできなかった。さらに、過去のチャットのログからの話題転換について、今回の実験では確認することができなかったが、音声認識誤りを話題にして会話が弾んでいる様子が見られたことから、STT 機能によるチャットのログが会話の話題に上る可能性は示された。

5.2 7 件法のアンケートの有意差があった項目について

有意差が見られた項目から ConvivialChat は Baseline システムよりチャットが見づらく使いにくいシステムだと判断されたものと考えられる。各機能の必要性を「はい/いいえ/わからない」で尋ねた質問の回答結果では、どの機能も「はい」と答えた回答者数を「いいえ」と答えた回答者数の方が上回っていた (図 4)。また、自由記述回答からは、Baseline システムの方を好むような意見が 2 名から得られた。この結果を受け、チャットの履歴が見やすく使いやすいシステムを目指した各機能の改良が必要であると考えられる。

ただ、この結果は、各機能が実験参加者にとって目新しく不慣れなものであったことによる評価であるとも考えられ、現状のシステムに利用者が慣れ親しむことで、より自然に使用され評価が変わる可能性も考えられる。実際に、機械音声の不自然さについて「音声が人工的なものであったために余計にメンバーの注意を引いたのではないか、ただ長時間使っていくうちに慣れてくると思うので時間が経てば何も感じなくなりそう」という意見を得ている。

6. まとめと今後の課題

本研究では、テキスト参加者とスピーチ参加者の双方が、その参加形態による疎外感や発言への抵抗に囚われることなく対等に話し合いを行うことができ、かつテキストとスピーチの相互補完的なコミュニケーションを可能にするシステムの提案を目的とし、TTS 機能、リアクションボタン読み上げ機能、STT 機能の 3 つの特徴的な機能を備えたシステム ConvivialChat を開発した。

さらに、ConvivialChat と Baseline システムを比較するコミュニケーション実験によってテキストとスピーチが共生するコミュニケーションシステムのデザインを探索した。その結果、TTS 機能には、スピーチ参加者と対等なテキスト参加者の話し合いへの参加を促す効果がある程度確認できた。また、リアクションボタンの読み上げ機能についても、テキスト参加者の相槌代わりとしての効果がある程度見込めることが示唆された。STT 機能については、議事録の効果は確認できたものの、話し合いへの参加具合をチャットのログから互いに対等に確認しあう効果、テキスト参加者の疎外感やメッセージ送信への抵抗を軽減する効果、過去のログを参照した話題転換を確認することはできな

かった。特に、スピーチ参加者の注意をチャットのログに向けることでテキスト参加者のメッセージの視認機会を向上させる効果については、STT 機能によるテキスト表示がチャットのログを見づらくしてしまう結果となり、期待したのとは全く逆の効果を得る可能性が示された。

それぞれの機能の今後の課題として、TTS 機能・リアクションボタン読み上げ機能については機械音声の質の向上、読み上げ方の工夫、加えてリアクションボタンの種類についてなど検討の余地が残っている。また、STT 機能については、音声認識の質の向上に加えて、この機能自体の必要性の再検討や目的の再設定の必要性が明らかになった。

さらに、ConvivialChat の今後の課題として、さらなる実験データの蓄積が必要であるほか、ビデオの有無と匿名性という 2 つの本研究では扱わなかった事項の検討、見やすさ・使いやすさの向上を目指したチャットのログのユーザインタフェース改善などが挙げられる。

参考文献

- [1] リモートミーティングでの音声認識の活用事例
https://github.com/DigitalNatureGroup/Remote_Voice_Recognition
- [2] Commentable <https://tetsuakibaba.github.io/commentable/>
- [3] Salvatore Andolina, Khalil Klouche, Diogo Cabral, Tuukka Ruotsalo, and Giulio Jacucci. InspirationWall: Supporting Idea Generation Through Automatic Information Exploration. In Proc. Creativity and Cognition 2015, pp.103-106, 2015.
- [4] Kenta Yamamoto, Ippei Suzuki, Akihisa Shitara, and Yoichi Ochiai. See-Through Captions: Real-Time Captioning on Transparent Display for Deaf and Hard-of-Hearing People. In arXiv e-prints. 5 pages, 2021.
- [5] 中西泰人, 倉持正之, 松川昌平: EnhancedChat: 音声と関連付け可能な 2 次元チャット, 情報処理学会論文誌, Vol. 43, No. 12, pp. 3550-3558, (2002).
- [6] 小倉加奈代, 西本一志: ChaTEL: マルチスレッド対話を容易にする音声コミュニケーションシステム, 情報処理学会論文誌, Vol. 47, No. 1, pp. 98-111 (2006).
- [7] 小倉加奈代, 西本一志, 小林智也, 杉山公造: 音声とテキストが混在するマルチモーダル・マルチスレッド・チャットシステムの試み, 情報処理学会研究報告, Vol. 2007-HCI-124, pp. 17-24, (2007).
- [8] Tscholl M., McCarthy J. D., Scholl J. The effect of videoaugmented chat on collaborative learning with cases, In Proc. CSCS 2005.