

博士学位論文

高齢者の発話促進を目的とした
発話単語画像を用いた傾聴システムに関する研究

令和6年3月

山崎 裕之

岡山県立大学大学院
情報系工学研究科

**高齢者の発話促進を目的とした
発話単語画像を用いた傾聴システムに関する研究**

目次

第1章 序論.....	1
1.1 本研究の背景.....	1
1.2 先行研究.....	2
1.3 本論文の目的.....	4
1.4 本論文の構成.....	5
参考文献.....	8
第2章 発話単語のうなずき画像を用いた傾聴システム.....	11
2.1 緒言.....	11
2.2 傾聴システムの開発.....	11
2.2.1 コンセプト.....	11
2.2.2 システム概要.....	12
2.2.3 提示画像.....	14
2.2.3.1 オフライン検索画像提示.....	14
2.2.3.2 オンライン検索画像提示.....	15
2.2.4 聞き手バーチャルエージェント.....	15
2.2.5 引き込み動作生成モデル.....	16
2.2.6 提示画像のうなずき動作.....	18
2.3 傾聴システムの評価.....	18

2.3.1	実験方法	18
2.3.2	実験結果	21
2.4	結言	25
	参考文献	26
第3章	ロボットの注視誘導機能	27
3.1	緒言	27
3.2	ロボットの注視誘導を用いた傾聴システムの開発	28
3.2.1	コンセプト	28
3.2.2	システム概要	29
3.2.3	聞き手ロボット	30
3.2.4	オンライン検索画像提示	31
3.3	ロボットの注視誘導を用いた傾聴システムの評価	32
3.3.1	実験方法	32
3.3.2	実験結果	35
3.4	結言	40
	参考文献	42
第4章	ロボットによるバックトラッキング機能	43
4.1	緒言	43
4.2	バックトラッキング機能	44
4.3	ロボットによるバックトラッキングを付与した傾聴システムの開発	44
4.3.1	コンセプト	44
4.3.2	ハードウェア構成	47
4.3.3	システム構成	48
4.4	ロボットによるバックトラッキングを付与した傾聴システムの評価	49
4.4.1	実験方法	49
4.4.2	実験結果	51
4.5	結言	55

参考文献.....	56
第5章 高齢者特性に合わせた統合システム	57
5.1 緒言.....	57
5.2 統合システムの開発.....	58
5.2.1 ハードウェア構成	58
5.2.2 システム構成.....	59
5.2.3 画像提示枚数.....	60
5.3 モニタリング調査	61
5.4 結言.....	63
第6章 結論	65
6.1 本研究のまとめ	65
6.2 今後の展望.....	68
謝辞.....	69
本論文に関する研究業績.....	70

第1章 序論

1.1 本研究の背景

我が国においては、2007年に高齢化率が21.5%となり「超高齢社会」を迎え、2022年には29.0%に達している。そのため、全世代が参画した、豊かな人生を享受できる超高齢社会の実現を目指す必要があるとの認識が示されている[1.1]。高齢者が豊かな人生を享受するためには、健康な生活習慣の維持や心の健康の維持、社会的なつながりの維持が求められる。そのような中で、高齢者にとって「話す」という行為は、身体機能維持に特に重要である[1.2]。しかし、介護者の人手不足などの要因により、高齢者の「話す」機会は失われていく一方である。高齢者介護施設においては、利用者同士が持参した写真やぬり絵の題材、テレビの画像、絵画等のような画像を見ることをきっかけとして話が開始されることがよくある。例えば、高齢者介護施設のレクリエーションでは、自分の思い出話をする際、何もない状態で話をするのと、画像を見ながら話をするのとでは、話の盛り上がりには大きな差が生じる。画像を見ながら話すと、話し手は当時の状況をより鮮明に、聞き手はその話をより具体的に想起することができる。また、お互いに同調し合い場が盛り上がることで話が次々と展開されていく。つまり画像を見ることで、自己の経験を想起させることにより、人々の発話を促進する働きがあるといえる。

同様に、高齢者の「話す」という行為を引き出すためには、聞き手の聞き方に依存する部分が多い[1.3]。具体的には、うなずきやあいづちを交えた傾聴であり、高齢者介護施設においては、介護従事者が主にその役割を担っている。特にうなずきを交えた傾聴は高齢者にとって単純で分かりやすい。しかし、人手不足が深刻な高齢者介護施設では、傾聴するための時間が取れないという問題点がある。このようなケースへの対策として、高齢者に対していつでも対話相手になれるような対話システムの必要性は高いと考える。

1.2 先行研究

画像と傾聴に関する研究は、これまでも行われてきた。画像を用いて高齢者の発話を促す方法としては、「回想法」と「共想法」が知られている。「回想」とは、人生の折々の経験や出来事がごく自然に思い出される心理的過程のことであり、1960年代初めにバトラー（Butler,R.N.）がそれまで否定的に捉えられてきた高齢者の過去への回想の価値を見直したのが回想法といわれるものである [1.4]。高齢者の人生に焦点をあて、過去、現在、未来へと連なるライフヒストリーを傾聴することを通じ、その心を支えることを目的とする技法である [1.5]。回想法では、語り手がどのような回想を語ったかという以上に、言葉の背景にある思いと、どのように語られるのか、という点を大切にしている [1.6]。回想法において、高齢者は知恵や経験の伝承者・発信者としての役割を果たし、受け手が心の底から敬意の念をもって真摯に受け止める時、高齢者の自尊心の回復や自己評価の向上に幾ばくか貢献できる可能性が生まれるとも言われている [1.5]。介護現場において、回想法は、認知症を含む高齢者に対する支援としてポピュラーな手法であり、対象者と実施者が一対一の関係においてなされる場合もあれば、対象者が複数のグループとなり、互いに思い出を語り合う場合もある。森田らは、過去を思い出すことと未来を展望することは裏表の関係にあり、適切な過去の回想は未来への活動に繋がると仮定し、個人化された認知モデルに基づく回想法である「モデルベース回想法」を提唱している [1.7]。その一環として、写真に対する高齢者の反応を予測しつつ、記憶を活性化させ、気力を充実させ、活動的な社会生活へ誘導する写真を提示していく取り組み [1.8]や、認知モデルによるモニタリングとケアを組み入れた高齢者支援システムの構築を行っている [1.9]。また、いくつになっても脳が健康な「生涯健康脳」を保つために、「知的好奇心」「運動」「コミュニケーション」の3つの要素と回想を組み合わせる方法も提唱されている [1.10]。

回想法が過去の経験を語るのに対して、共想法では現在の視点が語られる点において異なっており、双方向の活発な会話を行うことができるという。共想法は現在を中心に過去から未来まで多様なテーマを設定することで、常に新しい話題を見つける習慣を身に付けることを目指すものであり、こうした点を踏まえて共想法談話の性質を脱文脈化の観点から分析する取り組み [1.11] や、在宅高齢者の健康状態を対話によ

り観察するロボットシステムの開発 [1.12] などが行われている。大武は、テーマに沿った写真などの素材と話題を持ち寄り、時間を決めて話し手と聞き手が交代しながら会話し、想いを共有するという共想法を提唱している [1.13]。

画像を用いた研究としては、長野らによる、発話者の音声記録の結果から、文字や画像を提示することで発話内容の把握を支援するシステムがある [1.14]。瀬島らの研究では、身体引き込み反応する絵画オブジェクトを用いて、コミュニケーション支援への有効性を示している [1.15]。山崎・仲谷は、認知症患者と家族介護者とのコミュニケーション支援として写真を用いた思い出想起を活用し、家族との語り合いを支援するシステムを開発している [1.16]。また Tokunaga らは、高齢者の認知機能の低下を予防するために、思い出の写真を活用した対話型認知訓練システムを開発している [1.17]。

一方、傾聴のような聞き役対話システムを扱った研究では、単に言葉によるバーバル情報だけでなく、うなずきなどの言葉によらないノンバーバル情報を用いて対話者が相互に引き込み合い、身体性が共有される身体引き込み反応によって、円滑なコミュニケーションが行われている [1.18][1.19]。このような効果を持つ聞き手のうなずきやあいづちの反応は、ノンバーバルコミュニケーションの典型的な調整子であり、会話の流れを制御するだけでなく、身体性の共有により発話しやすくすることがわかっている。中でも、うなずきやあいづちが発想数に与える影響について、会議形式での議論において司会者が発言者の発言に対しうなずきやあいづちを行うことで発言者の発想数や想起する概念の数が増加する傾向にあることを示している [1.20]。ユーザの話を傾聴するようなシステムにおいては、文脈に応じて適切に多様な形態のあいづちをうつ必要があるとし、そのようなあいづちの生成を目標として、あいづち形態の分析・予測・生成に取り組んでいる研究もある [1.21]。これまでに著者らは、身体引き込みに着目し、発話音声のリズムに基づいてうなずきや身振り・手振りなど聞き手・話し手の身体引き込み反応動作を自動生成する技術 iRT (InterRobot Technology)を用いて、音声駆動型身体引き込み CG キャラクタ InterActor や音声駆動型身体引き込みロボット InterRobot を開発し、その有効性を示している [1.22]。

このように近年では、画像や傾聴システムを用いた対話システムの開発・研究が盛んに行われている。著者らも聞き手システムに傾聴効果を高めるために、発話単語の身体引き込みオブジェクトを用いた思い出想起促進システムを開発し、有効性を示している [1.23]。しかしながら、HMD (Head Mounted Display) を用いた没入型仮想空間において事前に作成された 3D オブジェクトを出現させるという限定的なシステム構成であるため、多様な発話に対応するためには、発話単語に対応するオブジェクトを事前に多数制作する必要があるという問題点がある。また、高齢者の使用を想定した場合、HMD の装着による負担が大きくなるという問題点も考えられる。

1.3 本研究の目的

本研究では、この画像と傾聴という 2 つのアプローチを組み合わせたシステムを提案する。具体的には、使用者の発話内容を音声認識し、発話内の単語に紐づく画像を画面上に提示し、その画像と、聞き手となるバーチャルエージェントがうなづく機能を付与した発話単語画像を用いた傾聴システムを開発、評価している。さらに傾聴効果を高めるために、ロボットを用いて実体感ある傾聴を表現することやロボットが使用者の話を聞き返すようにおうむ返しをするバックトラッキング技術を利用し確実に聞いてくれている安心感をフィードバックすることで、高齢者の発話促進を支援している。

傾聴システムの開発において、バーチャルエージェントは、実体を持つロボットに比べて低いコストで作成でき、また制御も容易である。一方で、場合によってはロボットを用いた方がユーザにとって話しやすいと主張されることがある [1.24]。ペットロボット Phyno を用いた音声対話システムの研究では、親しみやすさにおけるロボットの身体性による利点が主張されている [1.25]。ロボットは対話者と同じ空間を共有ことができ、実際に手で触れようと思えば触れられるという実在感が実体としての人の引き込みによる身体性の共有を飛躍的に高め、バーチャルエージェントより一体感や存在感を対話者に実感させることができる [1.26]。人間とロボットとの対話においては、「大きく反応すること」で相手に対して「話を聞いてもらえるという実感を与える」こと、ロボットの大きな振る舞いに気付いていなくても、人間は

対話相手の動きから対話相手に抱く印象が変わる可能性があることが示されている [1.27]. また, より自然な傾聴対話を行うために, 掘り下げ質問, 繰り返し応答, 語彙的応答, 評価応答という聞き手応答に加えて, 自分自身に関する話をする自分語りの生成を行うシステム [1.28] や, ユーザの感情評価と事前に設定した文章群を用いて発話文章を生成し, 連続した沈黙を認識した場合はシステムが直前に発話した文章に続く文章を生成するという沈黙認識に基づく音声対話システム [1.29] といった取り組みがある.

ロボットの特性やうなずき反応の技術により, 一定の親しみや話しやすさを感じる一方でバーバルな反応を求める意見が多く見られる. ロボットのような機器はあくまでも補助的なものであって, そこに対人援助技術や臨床的な支援の観点は必ず必要となると言われている [1.30]. ロボットによる傾聴をより好ましいものにするために考えられる技術として, 心理学の分野で用いられる相手とのコミュニケーション内容を相手にそのままフィードバックするバックトラッキング技術がある. 画像提示に使用する発話単語をバックトラッキングすることで, ロボットとの聞いてくれているという信頼関係を構築し, 安心感を与え, 自分の話を続けやすくして発話の促進を支援することが可能となる.

1.4 本論文の構成

本論文は本章を含め 6 章で構成されている. 本章を除いた 2 章以降の概要を以下に述べる.

2 章では, 音声から豊かなコミュニケーション動作を自動生成する技術を 3D バーチャルエージェントに組み込んだ **InterActor** を適用した発話単語のうなずき画像を用いた傾聴システムの開発, 開発したシステムを用いた高齢者介護施設の介護従事者を対象とした実験と評価について述べる. 発話単語のうなずき画像を用いた傾聴システムは, 高齢者介護施設に通う高齢者の施設での余暇時間を想定し, 聞き手バーチャルエージェントと画像オブジェクトを製作し, それらに身体引き込み反応を実装させたシステムである. 使用者の発話内容を音声認識し, 発話内の単語に紐づく画像を画面上に提示し, その画像と聞き手バーチャルエージェントがうなずくことにより, 発

話促進を目的とする。高齢者介護施設で高齢者を対象として実験を行うにあたり、まずは日頃から高齢者と触れ合っている介護従事者を対象に実験を行い、その知見を得ている。一対比較評価、官能評価ともに発話単語に基づく画像提示と提示画像にうなずき動作を付与したモードが肯定的に評価された。また、高齢者介護施設の現場で使えそうであるとの意見を得ており、今後の高齢者の新たな話し相手となりうる可能性が示されている。

3章では、実在感ある聞き手ロボットが提示画像へ注視誘導を行い、その画像と聞き手ロボットがうなずくことにより、発話を促進させることを目的とする傾聴システムを提案し、その有効性を検証している。常に画面を見続けることは注意力・集中力が低下した高齢者にとって困難であり、発話に基づく提示画像が注目されないといった課題を人型コミュニケーションロボットを聞き手に用いることで解決した。聞き手バーチャルエージェントと聞き手ロボットを用いて、高齢者介護施設に通う高齢者で実験を行い、そのシステムの有効性を示している。

4章では、使用者の発話内容を音声認識し、発話内の単語に紐づく画像を画面上に提示すると同時に聞き手ロボットが発話単語を音声によりバックトラッキングすることにより、発話を促進させることを目的とする傾聴システムを提案し、その有効性を検証している。2章の発話単語のうなずき画像を用いた傾聴システムに重要な要素として、聞き手が使用者の発話内容を聞いて理解しているかどうかといった信頼感に注目し、使用者が自身の話した内容が聞き手ロボットにどう伝わっているのかを確認することができるように聞き手ロボットにバックトラッキング機能を付与した。高齢者介護施設に通う高齢者で実験を行い、そのシステムの有効性を示している。

5章では、2章で開発したシステムをベースに3章、4章にて有効性を検証したシステムの機能を合わせた統合システムの開発について述べる。3章、4章の実験において一定数の画像が小さく見えにくいといった様子や意見が見られた。統合システムでは、高齢者の視力、認知力の低下を考慮し、提示画像枚数を減らすことで、高齢者の見やすさや分かりやすさを重視した改善を行った。高齢者介護施設の一角に統合システムを設置し、高齢者介護施設に通う高齢者にモニタリング調査を行い、そのシステムの有効性を示している。

6章では、本研究を通じて得られた成果を総括するとともに、今後取り組むべき課題について述べている。

参考文献

- [1.1] 内閣府, 令和5年版高齢社会白書 (2023), available from
<https://www8.cao.go.jp/kourei/whitepaper/w-2023/html/zenbun/s1_1_1.html
>, (参照日 2024年1月15日).
- [1.2] 井上映子, 和野千枝子, 山田万希子, 大森直哉, 後藤武, 川久保悦子, 熊谷玲子,
高齢者に対する「発話」促進援助による嚥下機能, 生理機能, 心理的側面への効
果, 城西国際大学紀要, Vol. 27, pp. 1–16 (2019).
- [1.3] 大竹裕也, 萩原将文, 評価表現による印象推定と傾聴型対話システムへの応用,
日本知能情報ファジィ学会誌, Vol. 26, No. 2, pp. 617–626 (2014).
- [1.4] バーバラ・K・ハイト, バレット・S・ハイト, 『ライフレビュー入門 – 治療
的な聴き手となるために』, ミネルヴァ書房, (2016).
- [1.5] 黒川由紀子, 認知症と回想法, 金剛出版, p. 78 (2008).
- [1.6] 野村豊子, 『ケアの現場・地域で活用できる回想法実践事例集 – つながりの
場をつくる47の取り組み』, 中央法規出版株式会社, (2022).
- [1.7] 森田純哉, 平山高嗣, 間瀬健二, 山田和範, メンタルタイムトラベルを誘導する
モデルベース回想法, 情報処理学会 研究報告ユビキタスコンピューティングシス
テム, Vol. 2015-UBI-47, No. 16, pp. 1–6 (2015).
- [1.8] 森田純哉, 平山高嗣, 間瀬健二, 山田和範, 認知アーキテクチャを組み入れた写
真スライドショーの開発: 展望と課題, 人工知能学会全国大会論文集, Vol. 2M4-
NFC-04b-3, pp. 1–3 (2015).
- [1.9] 森田純哉, 平山高嗣, 間瀬健二, 山田和範, 認知モデリングに基づく高齢者支援
の試み, 人工知能学会研究会資料 先進的学習科学と工学研究会, Vol. 73, pp. 95–
100 (2015).
- [1.10] 瀧靖之, 『回想脳 – 脳が健康でいられる大切な習慣』, 青春出版社, (2021).
- [1.11] 田中弥生, 小磯花絵, 大武美保子, 脱文脈化の観点から見た共想法に基づく高
齢者談話の分析, 国立国語研究所論集, Vol. 22, pp. 137–155 (2022).

- [1.12]熊谷和実, 徳永清輝, 三宅徳久, 田村和弘, 水内郁夫, 大武美保子, 見守り声かけロボットに対する高齢者による応答と印象評価, 日本ロボット学会誌, Vol. 39, No. 9, pp. 866–869 (2021).
- [1.13]大武美保子, 認知症予防回復支援サービスの開発と忘却の科学—共想法により社会的交流の場を生成する会話支援サービス—, 人工知能学会論文誌, Vol. 24, No. 6, pp. 569–576 (2009).
- [1.14]長野優一郎, 吉野孝, 拡張現実感技術を用いた発話可視化システム MIERUKEN の開発, 情報処理学会研究報告, Vol. 2009-DPS-141, No. 20, pp. 1–8 (2009).
- [1.15]瀬島吉裕, 石井裕, 渡辺富夫, アバタコミュニケーション支援のための音声駆動型身体的引き込み絵画を用いた仮想観客システム, 日本機械学会論文集, Vol. 78, No. 786, DOI: 10.1299/kikaic.78.523 (2012).
- [1.16]山崎和紘, 仲谷善雄, 思い出を用いた認知症者と家族介護者間におけるコミュニケーション支援, 情報処理学会第 74 回全国大会講演論文集, Vol. 2012, No. 1, pp. 229–230 (2012).
- [1.17]Tokunaga, S., Tamura, K. and Otake-Matsuura, M., A dialogue-based system with photo and storytelling for older adults: toward daily cognitive training, *Front. Robot. AI*, Vol. 8: 644964, DOI: 10.3389/frobt.2021.644964 (2021).
- [1.18]Watanabe, T., Human-entrained embodied interaction and communication technology, *Emotional Engineering*, Springer, pp. 161–177 (2011).
- [1.19]Morishima, S., *Non-verbal Communication System in Cyberspace, Usability Evaluation and Interface Design*, (2001).
- [1.20]大森晃, 土井晃一, あいづちが発想数に与える影響—その実験と分析—, 認知科学, Vol. 7, No. 4, pp. 292–302 (2000).
- [1.21]山口貴史, 井上昂治, 吉野幸一郎, 高梨克也, Nigel G.W., 河原達也, 傾聴対話システムのための言語情報と韻律情報に基づく多様な形態の相槌の生成, 人工知能学会論文誌, Vol. 31, No. 4, pp. C-G31_1–10 (2016).

- [1.22]Watanabe, T., Okubo, M., Nakashige, M. and Danbara, R., InterActor: speech driven embodied interactive actor, *International Journal of Human-Computer Interaction*, Vol. 17, No. 1, pp. 43–60 (2004).
- [1.23]窪田正勝, 渡辺富夫, 石井裕, 発話単語の身体引き込みオブジェクトを用いた思い出想起促進システムの開発, *ヒューマンインタフェース学会研究報告集*, Vol. 20, No. 3, pp. 193-196 (2018).
- [1.24]中野幹生, 実用的な対話ロボットの構築に向けてー物理世界での言語インタラクションのモデルと技術課題ー, *メディア教育研究*, Vol. 9, No. 1, pp. S29-S41 (2012).
- [1.25]玉川聡, 山本一公, 中川聖一, 音声対話インターフェースにおけるロボットエージェントの優位性の評価, *情報処理学会研究報告書*, Vol. 2013-SLP-97, No. 11(2013).
- [1.26]渡辺富夫, 身体的コミュニケーション技術とその応用, *システム制御情報学会誌*, Vol. 49, No. 11, pp. 431-436 (2005).
- [1.27]森木海翔, 楊潔, 菊池浩史, 菊池英明, ソーシャルワーカーの非言語行動スキルの対話ロボットへの実装に向けた検討, *人工知能学会 全国大会論文集*, Vol. JSAI2021, pp. 1-4 (2021).
- [1.28]石田真也, 井上昂治, 中村静, 高梨克也, 河原達也, 共感表出と発話促進のための聞き手応答を生成する傾聴対話システム, *人工知能学会 研究会資料 言語・音声理解と対話処理研究会*, Vol. 82, pp. 7-12 (2018).
- [1.29]前土佐勇仁, 三枝亮, 沈黙認識に基づく音声対話システムと感性評価, *情報処理学会 第84回全国大会講演論文集*, Vol. 2022, No. 1, pp. 949-950 (2022).
- [1.30]志村ゆず, 『レミニッセンス・セラピーー 回想法による高齢者支援』, 青灯社, (2022).

第2章 発話単語のうなずき画像を用いた傾聴システム

2.1 緒言

本章においては、使用者の発話内容を音声認識し、発話内の単語に紐づく画像を画面上に提示し、その画像と聞き手バーチャルエージェントがうなずくことにより、発話促進を目的とする傾聴システムを提案している。実装したプロトタイプシステムを用いて高齢者介護施設で実験を行い、その有効性を検証している。これにより、発話単語に応じて写真や絵画のオブジェクトが次々と提示され、それらが聞き手バーチャルエージェントと共にうなずき、積極的に関与して発話を促すことによって、高齢者介護施設での新たな対話相手となることが期待される。

2.2 傾聴システムの開発

2.2.1 コンセプト

本システムのコンセプトを図 2.1 に示す。本システムは、仮想空間内で発話音声に基づいて聞き手バーチャルエージェントの InterActor がうなずきによる身体引き込み反応をするとともに、発話に含まれる単語に基づく提示画像がうなずき反応をするシステムである。また、聞き手バーチャルエージェントは、画像が提示された方向に振り向き動作を行うことで、使用者の発話に対して傾聴意思があるかのように感じさせる。聞き手バーチャルエージェントだけでなく、自身の発話単語画像もうなずきによる身体引き込み反応をすることで、使用者とのリズム同調が增強され、使用者は自分の話を聞き流されずに受け止められている感覚が得られ、発話意欲が増大し発話が促進される。さらに、提示された発話単語画像は画面内に複数枚蓄積され、これまでの話した内容が確認できる。発話に伴う一連の提示画像は使用者の思い出想起や発想支援を加速させ、話題の展開力を豊かにし発話を支援する。高齢者にとっては、話す

ことは元気になる方法の一つである。画像から思い出を引き出し、発話を促進させる手法は、目新しい話題を手にすることが少ない高齢者に有効である。

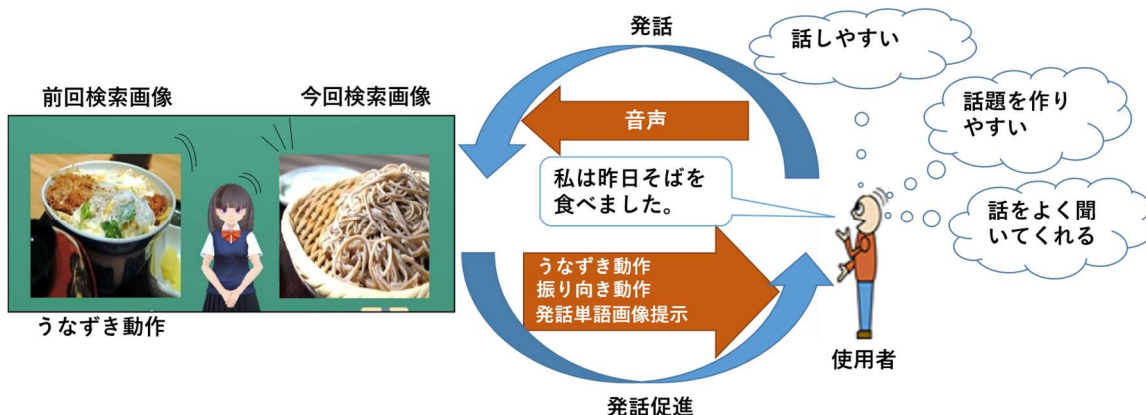


Fig. 2.1 発話単語のうなずき画像を用いた傾聴システムのコンセプト

2.2.2 システム概要

システムを使用する様子を図 2.2 に示す。本システムは、ノート PC (MSI GE66 RAIDER (CPU: Intel Corei9-10980HK 2.40GHz, Memory: 16GB, Graphics: NVIDIA GeForce RTX 2070 Super)), マイク (audio-technica AT9933USB) から構成される。画像はノート PC から出力しディスプレイ (DELL ST2410) に提示する。解像度は 1920×1080 pixel である。画面輝度は設定可能範囲の最大値である。画面色情報において、色形式は RGB を採用し、色空間には標準ダイナミックレンジを採用した。ディスプレイと使用者との距離は 900mm とした。本システムの環境構築には Unity2020.2.3 を用いた。使用者の発話単語の音声認識には Intel 社の RealSence SDK を用いた。本システムの使用画面を図 2.3 に示す。誰しも体験したであろう学生時分の思い出を想起するきっかけとなることを期待し、背景に黒板をイメージした画像を採用した。聞き手バーチャルエージェントは画面中央に配置しており、うなずきによる身体引き込み反応を示す。使用者の発話により画像が提示されたとき、聞き手バーチャルエージェントは画像の方向に振り向き反応をする。仮想空間上に画像は 10 枚まで提示でき、11 枚目からは 1 枚目の画像から逐次更新される。図

2.3 に示したシステムでは、画像を予めいくつか用意しておき、発話に応じて提示を行っている。画像についてはオンライン検索により提示することもできる。予め画像を用意する場合、使用者にとって関係の深い画像を用意することで、思い出の想起を促進することができる。一方、オンライン検索機能を用いると、使用者の予期しない画像が提示されることにより新たな発想やアイデアを生み出すきっかけにすることができる。



Fig. 2.2 使用風景

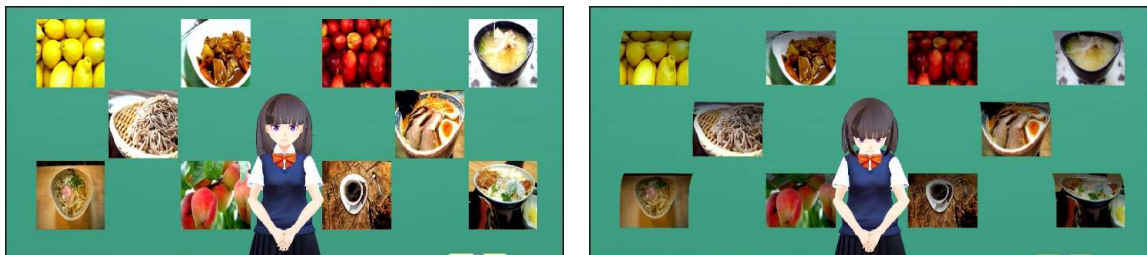


Fig. 2.3 使用画面 (左: 待機状態, 右: うなずき動作中)

2.2.3 提示画像

聞き手バーチャルエージェントとともに、使用者の発話を促進する機能として、提示画像がある。使用者が発話を行うと、対応する発話単語画像が画面に提示される。画像を提示するには、あらかじめローカル PC に用意した画像を使用するオフライン検索画像提示とインターネットより取得した画像を使用するオンライン検索画像提示の 2 種類の提示方法がある。本システムでは、オフライン画像提示方法を採用した。なお、画面内で一度に表示できる提示画像枚数の上限は 10 枚であり、上限以降は古い画像から順番に更新していく。提示位置は、あらかじめ設定した画像が重ならない 10 箇所である。以下に開発した画像検索機能について述べる。

2.2.3.1 オフライン検索画像提示

予め画像を PC 上に登録しておくことで、使用者の発話に対し画像を提示することができる。この方法では、候補となる画像と、タグとなる単語が 1 対 1 になるように予め手作業で登録しておく。使用者が発話を行うと、登録された画像の中から、音声認識によって得られた発話内の単語と一致するものを検索し、提示する。例えば、「スイカ」の画像を出したいとき、システムに予めスイカの画像と「スイカ」というワードを登録しておく。その後使用者が発話を行い、「昨日スイカを食べた」等の発話があったとき、音声認識結果より「スイカ」というワードが検出されるため、登録したスイカの画像を提示する。この方法では、手作業にて画像を登録することができるため、使用者にとって関係の深い画像や思い出のある写真なども登録することができる。そのため、より深い内容の発話を促進することができる。ただし、予め登録する必要があるため、登録単語以外の単語に対し画像を提示することはできない。オフライン検索の流れを図 2.4 に示す。

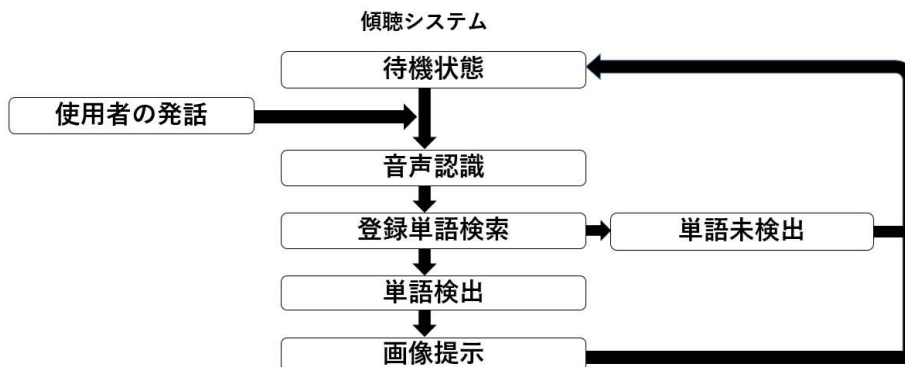


Fig. 2.4 オフライン検索フロー図

2.2.3.2 オンライン検索画像提示

オンライン検索機能を用いて画像を提示する。音声認識によって得られた発話文を形態素解析し、一般名詞を抽出する。抽出した名詞から画像検索 API により画像検索を行い、検索候補があればその第 1 候補から第 3 候補までの画像の中からランダムに 1 枚選択し、画像提示を行う。画像検索 API には PixabayAPI を用い、形態素解析には Mecab を用いる。この方法では、外部 API を用いるため、手作業により画像を登録するよりもはるかに多くの名詞に対応した画像を提示することができるが、固有名詞や、画像検索で見つからなかった場合は提示できない。しかし、登録された画像ではないことから、使用者の予期しない画像が提示されることがあるため、様々な画像が提示されることにより、発想支援としての効果も期待できる。一方で、インターネット接続による使用が前提であること、音声認識から画像提示まで、3 秒程度の通信遅延によるタイムラグが生じることなどが問題点として挙げられる。

2.2.4 聞き手バーチャルエージェント

使用者のより自然な発話を促すために、聞き手として人型の 3D バーチャルエージェントを採用した。聞き手バーチャルエージェントは、使用者と対面の位置関係になることを想定し、画面中央に配置した。聞き手バーチャルエージェントは、発話単語の提示に合わせて振り向くことで共同注視を行う。振り向き動作を行う様子を図 2.5

に示す。振り向き動作は、共同注視により視線の共有を行うことで、使用者が伝える事柄が正確に同定されることにより伝わっていると感じやすくするために導入した [2.1]。聞き手バーチャルエージェントは、画像が提示された方向に対し上下左右に振り向き動作をする。この動作により、聞き手バーチャルエージェントが使用者の発話に対し傾聴意思があるように感じさせることができ、また振り向いた方向に画像が提示されたことを使用者に共有認識させることができる。聞き手バーチャルエージェントが画像に視線を向け発話の情報を共有していることを示すことは、聞き手の理解の根拠となる有効かつノンバーバルなフィードバックとなる。

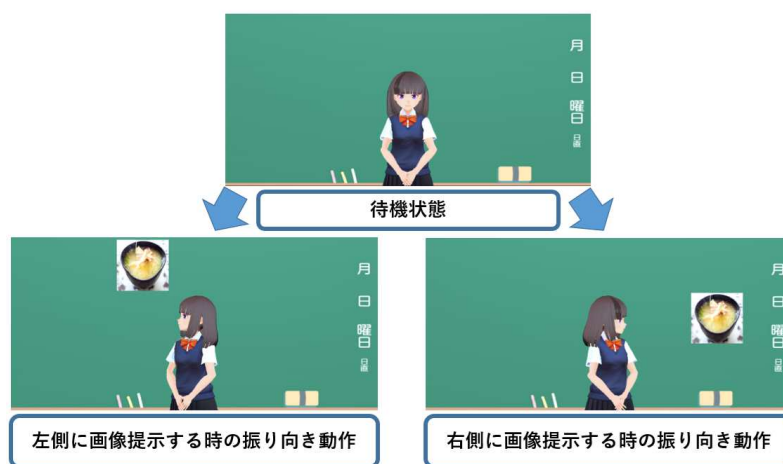


Fig. 2.5 聞き手バーチャルエージェントの新規提示画像への振り向き動作

2.2.5 引き込み動作生成モデル

インタラクションモデルの構築のために、人と人の対面コミュニケーションでの音声と身体動作の関係を分析した結果、胴部・腕部の動作は頭部の動作との関係が強く、うなずきと頭部の動きには強い相関があることから、聞き手の身体動作はうなずきを主体としていることが明らかになった [1.22]。その解析結果に基づき、聞き手バーチャルエージェントのインタラクションモデルとして、音声の ON-OFF パターンに基づくうなずき反応モデルを導入している(図 2.6)。

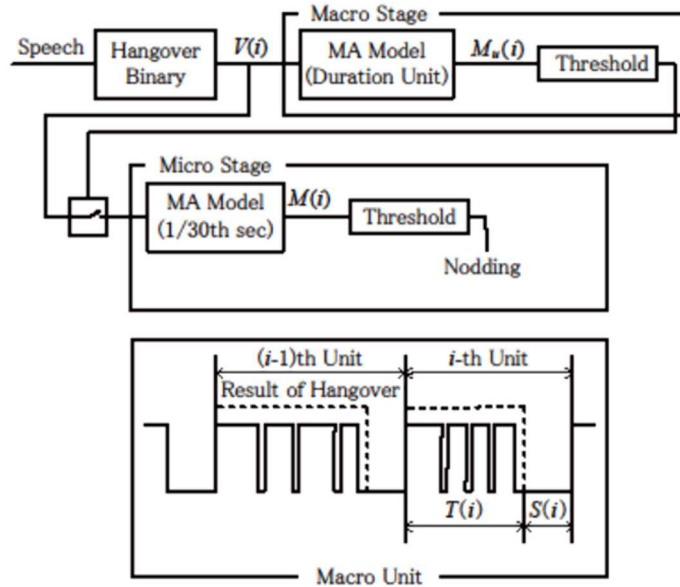


Fig 2.6 うなずき反応モデル

うなずきの予測モデルはマクロ層とマイクロ層からなる階層モデルである. マクロ層では, 音声の呼気段落区分での ON-OFF 区間にうなずきの開始が存在するか $M_u(i)$ をユニット時間率 $R(i)$ (式(2))の線形結合で表される式(1)の MA(Moving-Average)モデルを用いて予測する. マイクロ層では, 式(3)を用いて MA モデルでうなずきの開始時点 $M(i)$ を推定する.

$$M_u(i) = \sum_{j=1}^J a(j)R(i-j) + u(i) \quad (1)$$

$$R(i) = \frac{T(i)}{T(i) + S(i)} \quad (2)$$

$$M(i) = \sum_{k=1}^K b(k)V(i-k) + w(i) \quad (3)$$

$a(j)$: Prediction coefficient

$T(i)$: ON section in the i -th unit

$S(i)$: OFF section in the i -th unit

$u(i)$: noise

$b(k)$: Prediction coefficient

$V(i)$: Voice data

$w(i)$: noise

2.2.6 提示画像のうなずき動作

本システムでは, 聞き手バーチャルエージェントと提示画像がうなずき反応を行う. 提示画像のうなずきの様子を図 2.7 に示す. 提示画像の上部 1/3 を回転させることで, 人の頭部動作を表現する. 複数枚画像が提示されたときは, 使用者の発話に対しすべての提示画像が同じタイミングでうなずき動作を行う. 提示位置でうなずき動作を真正面にさせると, 端にある提示画像がそっぽを向いているように感じるため, 使用者に対してうなずき動作を行うように調整する. 提示画像全てにうなずき反応を付与しているが, 最新提示画像のみにうなずき反応を付与する手法も考えられる.



Fig 2.7 画像のうなずき反応(左:待機,右:うなずき)

2.3 傾聴システムの評価

2.3.1 実験方法

振り向き動作を付加した聞き手バーチャルエージェントとともに, 表 2.1 に示す 4 種類のモードを用意し, 評価実験を行った. A モードは提示画像のうなずき無し+ランダム画像提示, B モードは提示画像のうなずき有り+ランダム画像, C モードは提示画像のうなずき無し+発話単語画像提示, D モードは提示画像のうなずき有り+発話単語画像提示で実験を行った. 実験システムの概略を図 2.8 に示す. A モードと B モードの「ランダム画像」とは, 発話単語とは無関係の画像をランダムに提示するというものである. 約 200 個の画像の中から, 乱数発生アルゴリズムによってランダムに選択している. 提示位置は発話単語画像と同様にあらかじめ設定した 10 箇所をランダムに提示される. C モードと D モードの「発話単語画像」は, 2.2.3.1 節で述

べたオフライン検索方式により、AモードとBモードと同様の約200個の画像の中から、発話内の単語に紐づくものを選択している。2.2.3.2節で述べたオンライン検索方式は、画像検索の精度が十分に検証できておらず、また提示までにタイムラグが生じるという問題点が未解決のため、今回の実験では使用していない。また、高齢者介護施設で実験を行うにあたり、介護従事者に事前ヒアリングを行った結果、要介護者である高齢者に直接システムに触れてもらうのは現状難しいという意見が得られたため、まずは介護従事者を対象に実験を行い、その知見を得ることとした。実験手順は、まず実験概要を実験参加者に説明し、各モードを試行して反応の様子を確認させた。発話内容は食べ物に関する自由発話とし、例えば、「昨日の晩ご飯は～を食べた」、「好きな食べ物は～だ」、「～が美味しかった」といった内容の話をしてもらうよう説明した。その後、実験参加者に対して4種類のモードから2種類のモードを抽出し、2種類のモードに対しそれぞれ1分間体感させ、一対比較させた。一対比較は実験参加者1人につき6(=4C2)通り行った。次に、各モードを1分間体感させ、表2.2の8項目について7段階(中立0)で官能評価させた。同時に各モードについて自由記述をさせた。実験参加者は、20～76歳の介護従事者24名であり、性別は男性9名、女性15名、年代別は20代5名、30代4名、40代5名、50代5名、60代以上5名であった。なお、本実験は岡山県立大学倫理委員会の承認を得た上で、事前に実験参加者に目的・方法を口頭と文章にて説明し、同意を得て実施した。録画・録音されたデータは研究者以外の目に触れないこと、匿名性の保護、研究への参加を拒否した場合も不利益を受けないこと、得られたデータを研究以外の目的で使用しないこと等を保障した。

Table 2.1 システム評価実験条件.

モード	画像のうなずきの有無	提示画像の種類
Aモード	うなずき無	ランダム画像
Bモード	うなずき有	ランダム画像
Cモード	うなずき無	発話単語画像
Dモード	うなずき有	発話単語画像

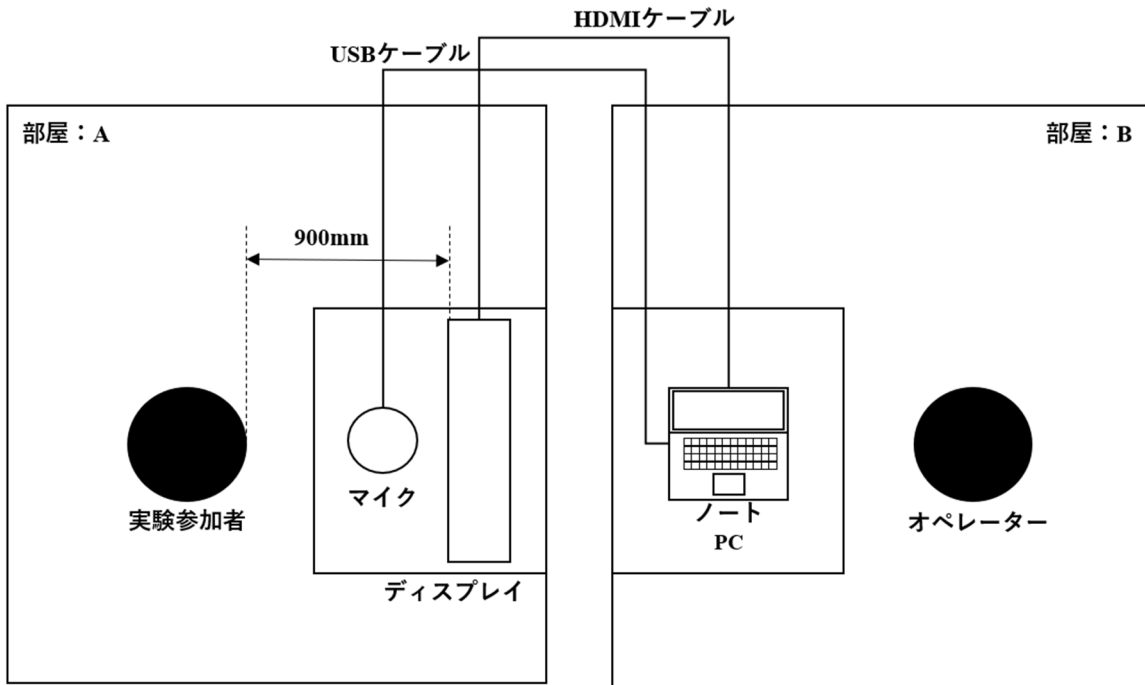


Fig. 2.8 実験システムの概略

Table 2.2 アンケート項目

No.	項目
①	楽しみを感じましたか
②	親しみを感じましたか
③	システムをまた利用したいですか
④	話しやすかったですか
⑤	1分より短く感じましたか
⑥	話が伝わったと感じましたか
⑦	話そうとする気持ちが高まりましたか
⑧	話題が頭に浮かびましたか

2.3.2 実験結果

4種類のモードに対する一対比較の結果を表 2.3 に示す。表中の数字は、各列に対し各行の表示をより好ましいと答えた実験参加者の数を表している。この一対比較の結果に対して、Bradley-Terry モデル ($P_{ij} = \pi_i / (\pi_i + \pi_j)$, $\sum \pi_i = \text{const.} (= 100)$, $\pi_i: i$ の強さ, $P_{ij}: i$ が j に勝つ確率) を想定し、各モードの強さ π を最尤推定した結果を図 2.9 に示す。これにより、画像のうなずき有で発話単語画像を提示する D モードが最も高く評価されており、続いて C モード, B モード, A モードの順に評価されたことが分かる。

Table 2.3 一対比較の結果.

	A	B	C	D	合計
A		9	6	3	18
B	15		5	6	26
C	18	19		8	45
D	21	18	16		55

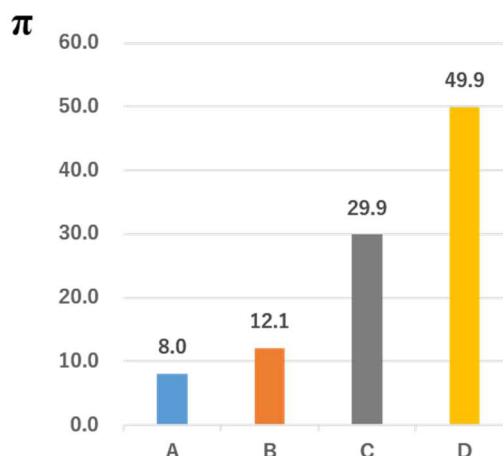
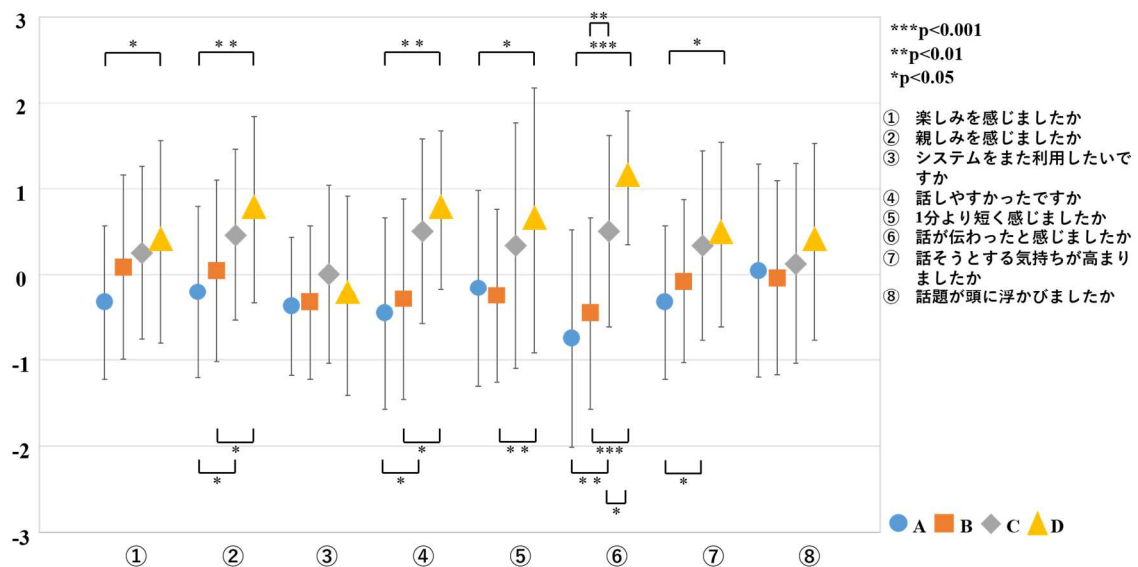


Fig. 2.9 Bradley-Terry モデルに基づく π の結果

次に、7段階評価結果を図 2.10 に示す。図には各モードの平均値とその標準偏差を示している。図より全体的に一対比較実験と同様の評価傾向があることがわかる。Friedman の分散分析および Wilcoxon の符号順位検定を用いた多重比較を行った。その結果、A モードと D モードの間で、「① 楽しさを感じましたか」、「⑤ 1分より短く感じましたか」、「⑦ 話そうとする気持ちが高まりましたか」の項目で有意水準 5%、「② 親しみを感じましたか」、「④ 話しやすかったですか」の項目で有意水準 1%、「⑥ 話が伝わったと感じましたか」の項目で有意水準 0.1%、A モードと C モードの間で、「② 親しみを感じましたか」、「④ 話しやすかったですか」、「⑦ 話そうとする気持ちが高まりましたか」の項目で有意水準 5%、「⑥ 話が伝わ

ったと感じましたか」の項目で有意水準 1%，Bモードと Dモードの間で、「② 親しみを感じましたか」，「④ 話しやすかったですか」の項目で有意水準 5%，「⑤ 1分より短く感じましたか」の項目で有意水準 1%，「⑥ 話が伝わったと感じましたか」の項目で有意水準 0.1%，Bモードと Cモードの間で、「⑥ 話が伝わったと感じましたか」の項目で有意水準 1%，Cモードと Dモードの間で、「⑥ 話が伝わったと感じましたか」の項目で有意水準 5%で有意差が認められた。「③ システムをまた利用したいですか」および「⑧ 話題が頭に浮かびましたか」の項目ではどのモード間においても有意差が認められなかった。



	A			B		C
	B	C	D	C	D	D
①	-	-	5%	-	-	-
②	-	5%	1%	-	5%	-
③	-	-	-	-	-	-
④	-	5%	1%	-	5%	-
⑤	-	-	5%	-	1%	-
⑥	-	1%	0.1%	1%	0.1%	5%
⑦	-	5%	5%	-	-	-
⑧	-	-	-	-	-	-

Fig. 2.10 アンケートによる官能評価の結果

また、7段階評価結果を二要因分散分析した結果を表 2.4 に示す。うなずきの有無に関する画像群と発話単語画像の有無に関する群に分類したものを二要因分散分析した結果、発話単語画像の群で「① 楽しさを感じましたか」、「⑤ 1分より短く感じましたか」の項目で有意水準 5%、「② 親しみを感じましたか」、「⑦ 話そうとする気持ちが高まりましたか」の項目で有意水準 1%、「④ 話しやすかったですか」、「⑥ 話が伝わったと感じましたか」の項目で有意水準 0.1%で有意差が認められた。また、画像うなずきの群で「⑥ 話が伝わったと感じましたか」の項目で有意水準 5%で有意差が認められた。二要因分散分析の結果、すべての項目で交互作用がみられず、因子：発話単語画像は「③ システムをまた利用したいですか」と「⑧ 話題が頭に浮かびましたか」以外の項目で主効果が認められた。因子：画像うなずきは「⑥ 話が伝わったと感じましたか」でのみ主効果が認められた。このことより、提示された発話単語画像および画像うなずきの効果により、使用者が話しやすくなる発話促進効果があることを示した。

Table 2.4 二要因分散分析の結果

項目\因子	発話単語画像	画像うなずき	発話単語画像 &画像うなずき
①	0.048*	0.218	0.505
②	0.002 **	0.213	0.923
③	0.268	0.614	0.480
④	0.000 ***	0.356	0.853
⑤	0.013 *	0.702	0.492
⑥	0.000 ***	0.045 *	0.463
⑦	0.005 **	0.376	0.767
⑧	0.310	0.734	0.497

*p<0.05 **p<0.01 ***p<0.001

なお、自由記述形式のアンケート調査結果によると、CモードとDモードの自由記述では、「話した画像が出てくると楽しい」「話の内容の画像が映し出されたのでうれしかった」「自分が話した内容の画像が出ていたので、話しやすく感じた」などの発話単語画像に対するポジティブな意見が多く見られた。一方で「関連画像が上手く提示される時とされない時があり、話が伝わっているように感じなかった」といった意見があり、登録されていない発話がされたと考えるが、提示されると期待しているところで提示されないとネガティブなイメージに繋がるということが確認された。Dモードの自由記述では、「話に合わせて画像がうなずくので話しやすかった」「うなずくので話が伝わったと感じた」「背景の画像が同調して、話しに集中することができた」などのうなずきに対する意見も見られた。AモードとBモードの自由記述では、「発話内容と違う画像が出てくることが多く、上手く伝わっていないと感じた」「話の内容と違う画像だったので、つまらなかった」「画像が出たが会話内容と違う画像だったので話しにくかった」などのランダム画像に対するネガティブな意見が多く見られた。しかし、Bモードの自由記述では、「背景の画像が同調してくれるので親しみや話しやすさや傾聴感が増した」などのうなずきに対するポジティブな意見も見られた。BモードとDモードに対する記入の中で、両モードともに画像の動きをうっとうしく感じるといったネガティブな意見と同調してくれて話しやすいと感じるといったポジティブな意見が見られた。Dモードでは、同調してくれて話しやすいと感じる意見が多く見られ、Bモードでは、画像の動きをうっとうしく感じる意見が多く見られた。発話以外の画像が出たことで余計に違和感を持つことや提示画像を絵として認識した故の現実では起こりえない事象に違和感を持つなどといった、実験参加者毎の提示画像に対する捉え方の違いが見られた。また、高齢者の使用を想定とした感想として「現場で使ってみたい」、「高齢者に使用すると面白そう」、「喜ぶと思う」、「女の子に話しかけそう」等のポジティブな意見が見られた一方で「画面をずっと見るのは難しそう」、「画像が小さく見えない、気が付かない」、「長く使っていると疲れる」、「画面に話しかけない」等のネガティブな意見が見られた。日頃より高齢者と関わる介護従事者の経験による高齢者の注意力・集中力の低下や視力等の身体的状態を危惧した意見と見られる。

2.4 結言

本章では, 発話内容を音声認識して, 発話内の単語に紐づく画像を画面上に提示し, その提示画像がうなずくことにより身体引き込み反応を示す傾聴システムを開発した. 本システムは, 聞き手バーチャルエージェントに画像提示のタイミングで画像に振り向き動作を行わせ共同注視させることで, 使用者に聞き手バーチャルエージェントの傾聴意思を伝える機能を有している.

本システムを用いて, 高齢者介護施設において評価実験を行い, 官能評価により有効性を検証した. 検証の結果, 画像提示とうなずき動作による傾聴効果により, 使用者が話しやすくなる発話促進効果があることが確認できた. しかしながら, 発話単語画像が画面上に提示され, 聞き手バーチャルエージェントも画面上で誘導を行っているため常に画面を注視していなければならないなど関わり方に制限ができていた. これは「画面をずっと見るのは難しそう」, 「長く使っていると疲れる」といった意見にも見られるように, 高齢者にとって画面を常に注視し話を続けることは, 注意力や集中力の維持等の面で負担になると考える. 今回の実験は, 介護従事者を対象に行ったものであり, 高齢者介護施設に入居している高齢者に行ったものではなく, 実験時の話題も食べ物に関するものに限定しており, 提示画像もそれに対応した一定数のものに制限される. 今回の実験で発話内容の画像が提示された方が良いといった評価を得ていることから, 様々な発話に対して十分に対応する必要がある. これらに対応した上で高齢者を対象とし, 様々な発話に対する画像を制限することなく提示し, 有効性の検証を行う必要がある. 介護従事者を対象とした本評価実験において, 高齢者介護施設の現場で使えそうであるとの感想を得ており, 今後の高齢者の新たな話し相手となりうる可能性が示されたが, 目的である高齢者の発話を促進するという観点から本実験にて得られた知見を基に高齢者が使用する上で考慮すべき点を踏まえて, 本システムの発展を目指す.

参考文献

- [2.1]栗山直子, 寺井あすか, 安原正晃, 徳永健伸, 山岸侯彦, 楠見孝, 共同問題解決時の二者の視線一致が共有知識の形成に及ぼす効果, 日本認知科学会, Vol. 27, pp. 2-31 (2010).

第3章 ロボットの注視誘導機能

3.1 緒言

2章では、発話の促進を目的とした仮想空間内で発話音声に基づいて聞き手バーチャルエージェントのInterActorがうなずきによる身体引き込み反応をするとともに、発話に含まれる単語に基づく提示画像がうなずき反応をする傾聴システムを開発し、有効性を示した。しかし、画面に常に注視していなければならないなど関わり方に制限ができていた。これは自身の発話単語画像をみることで身体引き込み効果が得られるためであるが、常に画面を見続けることは注意力・集中力が低下した高齢者にとって困難である。また、発話単語に対応する画像を事前に取得しシステムに登録する必要があり手間を要することから、様々な発話に含まれる単語に十分に対応できていないといった課題が残された[3.1]。

本章では、使用者の発話内容を音声認識し、発話内の単語に紐づく画像をオンライン上で検索し画面上に提示し、同時に人型コミュニケーションロボットが、提示された画像に注視誘導を行う傾聴システムを提案する。実装したシステムを用いて、高齢者介護施設の利用者を対象として実験を行い、有効性の検証を行う。本システムは、使用者の発話単語に応じて画像のオブジェクトが画面上に提示されると同時に聞き手ロボットがその提示画像へ注視誘導することで画像が提示されたことを強調し認識しやすくする。聞き手バーチャルエージェントより存在感がある聞き手ロボットにより画像が提示されたことを実空間で明示的に表現することができる。それらの画像が聞き手ロボットと共いうなずき、積極的に関与して発話を促すことによって、高齢者介護施設での新たな対話相手となることが期待される。

3.2 ロボットの注視誘導を用いた傾聴システムの開発

3.2.1 コンセプト

本システムは, 実空間で発話音声に基づいて聞き手ロボットがうなずきによる身体引き込み反応をするとともに, 発話に含まれる単語に基づく提示画像へ聞き手ロボットの動作によって注視誘導を行う傾聴システムである(図 3.1). 使用者が発話を行うとその発話単語に応じた画像が画面に提示され, 聞き手ロボットはその画像の位置情報をもとに振り向きなどの身体動作を用いて注視誘導を行う. 聞き手に聞き手バーチャルエージェントではなく物理的な聞き手ロボットを用いることで, 物理的な存在感を与え, 傾聴をよりリアルで心地よいものに変え, 使用者の信頼感や共感を深める. 聞き手ロボットの存在が実際の空間に広がり, 使用者に現実とバーチャルを交互に行き来させながら使用者の注意と集中を一層高めることで, 使用者はよりリアルなコミュニケーションを享受できる. 聞き手ロボットだけでなく, 自身の発話単語画像もうなずきによる身体引き込み反応をすることで, 使用者とのリズム同調が增強され, 使用者は自分の話を聞き流されずに受け止められている感覚が得られ, 発話意欲が増大し発話が促進される. 存在感を伴う聞き手ロボットの振る舞いは, 使用者が感じとる聞き手ロボットの感情や状態を豊かに表現し, 使用者の発話に対して理解を示す物理的な根拠となり, 使用者に深い共感を与える. さらに, 聞き手ロボットはフィジカルな触れ合いを通じて, 人間と接しているような親密性を提供することができる. このように本システムでは, 聞き手にロボットを用いることで実在感のある対話を実現しつつ, 画像によるインタラクション効果を実空間で明示的に表現することができる.

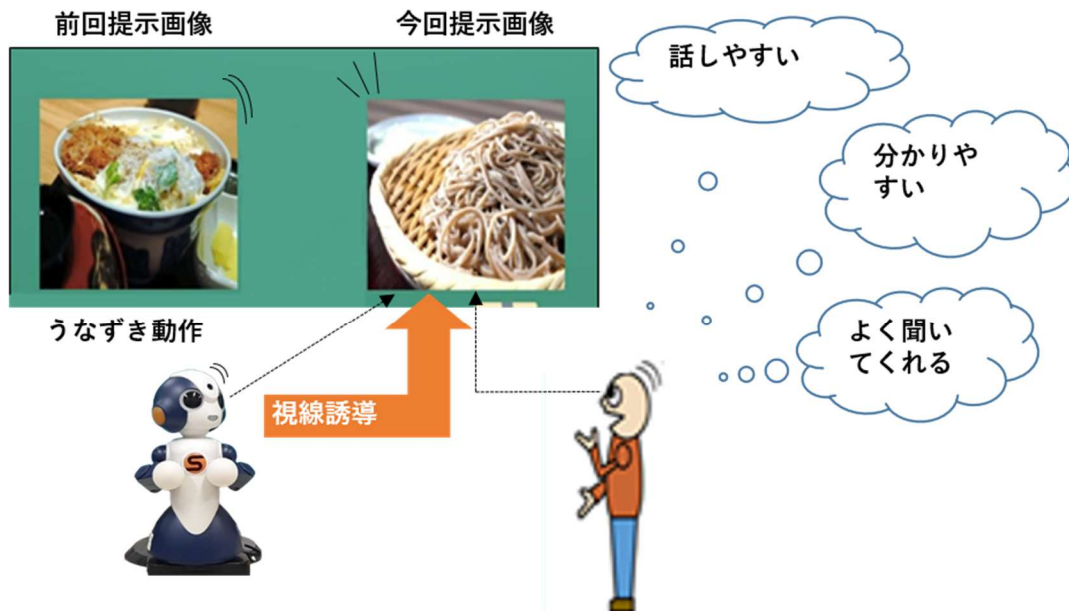


Fig 3.1 ロボットの視線誘導による発話音声のうなずき画像を用いた傾聴システムのコンセプト

3.2.2 システム概要

本システムは、画像描画用 PC (MSI GE66 RAIDER (CPU: Intel Corei9-10980HK 2.40GHz, Memory:16GB, Graphics: NVIDIA GeForce RTX 2070 Super)), 音声集音用マイク (audio-technica AT9933USB), 人型コミュニケーションロボット(ヴイストン社 Sota) から構成される。画像は画像描画用 PC から出力しディスプレイ (DELL ST2410) に提示する。解像度は 1920×1080 pixel である。画面輝度は設定可能範囲の最大値である。画面色情報において、色形式は RGB を採用し、色空間には標準ダイナミックレンジを採用した。本システムの環境構築には Unity2020.2.3 を、使用者の発話単語の音声認識には Intel 社の RealSense SDK を用いた。聞き手ロボットのシステム構築には統合開発環境 Eclipse にて Java を用いて開発を行い、ヴイストン社が発行しているライブラリを使用した。聞き手ロボットと画像のうなずきのインタラクションモデルとして、音声の ON-OFF パターンに基づくうなずき反応モデルを導入している。

先行システムと本システムの使用風景を図 3.2 に示す。聞き手ロボットは画面中央に配置し、うなずきによる身体引き込み反応を示す。使用者の発話により画像が提示

されたとき、聞き手ロボットは提示画像の方向に振り向き反応をする。仮想空間上に画像は10枚まで提示でき、11枚目からは1枚目の画像から逐次更新される。画像提示方式には、画像を予めいくつか用意しておくオフライン検索方式[3.1]ではなく、発話に応じて提示されるオンライン検索方式を採用した。



Fig. 3.2 使用風景 (左: 先行システム, 右: 本システム)

3.2.3 聞き手ロボット

聞き手ロボットには、画面サイズに合わせてヴイストン社の **Sota** を採用した。聞き手ロボットは、使用者と対面の位置関係になることを想定し、画面中央に配置し、うなずきによる身体引き込み反応を示す。うなずき動作を行う様子を図 3.3 に示す。また、画像提示が行われた方向の腕を振り上げ、首と体を上下左右に回転させる振り向き動作を行うことで視線を誘導し共同注視を行う。人間のコミュニケーションにおいて、相手が示している事柄が何を指すのかを正確に同定しお互いに情報を共有することが重要であると論じられている[2.1]。共同注視により視線の共有を行うことで、使用者が伝える事柄が正確に同定されることにより伝わっていると感じやすくなるのである。振り向き動作を行う様子を図 3.4 に示す。この動作により、使用者は自身の発話に対して聞き手ロボットが傾聴意思を持っているかのように感じることができる。また、実在感のある聞き手ロボットが大きな動きで画像が提示された方向へ注視誘導し共同注視をしているため、提示画像の情報を共有認識することができる。聞き手ロボットが画像に視線を向け発話の情報を共有していることを示すことは、聞き手ロボットの理解の根拠となる有効かつノンバーバルなフィードバックとなる。



Fig 3.3 聞き手ロボットのうなずき反応(左:待機,右:うなずき)

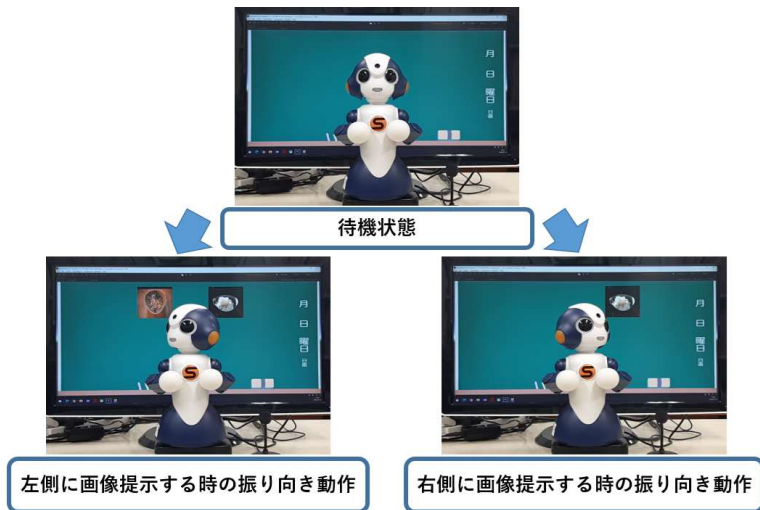


Fig. 3.4 聞き手ロボットの画像への視線誘導動作の遷移

3.2.4 オンライン検索画像提示

オンライン検索の流れを図 3.5 に示す. 聞き手とともに, 使用者の発話を促進する機能として, 画像提示がある. 使用者が発話を行うと, 対応する発話単語画像が画面に提示される. なお, 画面内で一度に表示できる提示画像枚数の上限は左右 5 枚ずつ計 10 枚であり, 上限以降は古い画像から順番に更新していく. 提示位置はあらかじめ設定した画像が重ならない 10 箇所である.

オンライン検索では, まず音声認識(Intel RealSense SDK)によって得られた発話を形態素解析し, 一般名詞を抽出する. 抽出した名詞から画像検索 API により画像検索を行い, 同じ発話単語の繰り返して同一画像が複数されるのを防ぐため, 検索候補があればその第 1 候補から第 3 候補までの画像の中からランダムに 1 枚選択し,

画像提示を行う。画像検索 API には PixabayAPI を用い、形態素解析には.NET 日本語形態素解析エンジン NMeCab を用いる。オンライン検索では、外部 API を用いるため、手作業により画像を登録するよりもはるかに多くの名詞に対応した画像を提示することができるが、固有名詞や、画像検索で該当画像が 1 枚も見つからなかった場合は提示できない。登録された画像ではないことから、使用者が思い描くイメージと違う画像が提示されることがあるため、聞いてくれていないや受け入れられていないと感じてしまうことも考えられるが、一方で様々な画像が提示される可能性があるため、発想支援としての効果も期待できる。

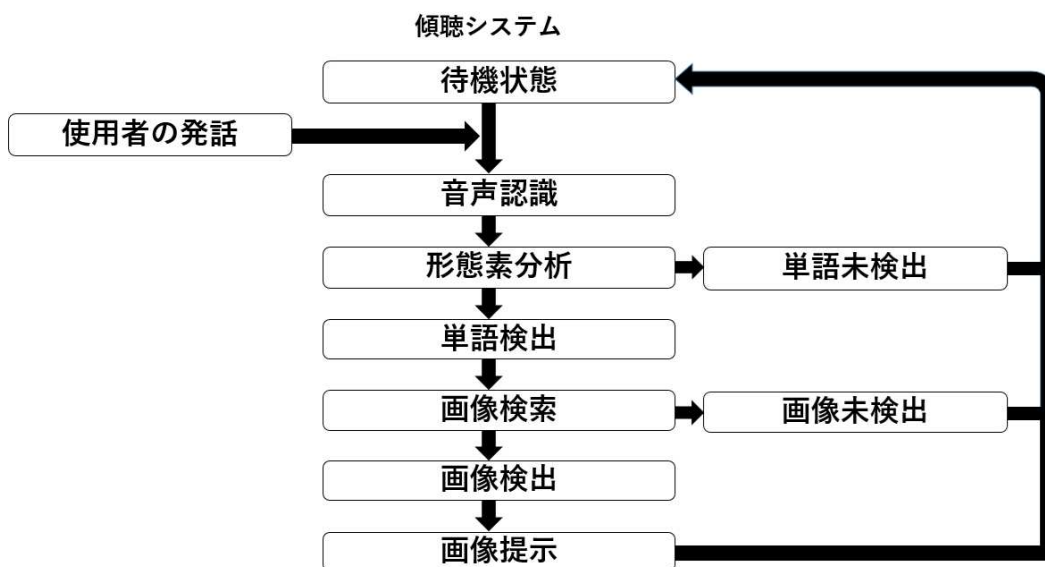


Fig. 3.5 オンライン検索フロー図

3.3 ロボットの注視誘導を用いた傾聴システムの評価

3.3.1 実験方法

新たに実装した視線誘導動作を付加した聞き手ロボットとともに、表 3.1 に示す 4 種類のモードを用意し、評価実験を行った。A モードは聞き手バーチャルエージェント+提示画像のうなずき無し、B モードは聞き手バーチャルエージェント+提示画像のうなずき有り、C モードは聞き手ロボット+提示画像のうなずき無し、D モードは聞き手ロボット+提示画像のうなずき有りで行った。実験システムの概略を図

3.6 に示す．すべてのモードにおいて画像提示手法としてオンライン検索方式を用いた．提示位置はあらかじめ設定した 10 箇所にランダムに提示される．A モードと B モードの聞き手バーチャルエージェントには，人型の 3D バーチャルエージェントを採用した．C モードと D モードの聞き手ロボットには，人型コミュニケーションロボットの Sota を採用した．聞き手バーチャルエージェントと聞き手ロボットどちらも，使用者と対面の位置関係になることを想定し，画面中央に配置した(図 3.2)．また，発話単語の提示された方向に合わせて振り向くことで共同注視を行う．

実験手順は，まず実験概要を実験参加者に説明し，各モードを使用直前に試行して反応の様子を確認させた．発話内容は食べ物に関する自由発話とし，例えば，「昨日の晩ご飯は～を食べた」，「好きな食べ物は～だ」，「～が美味しかった」「得意料理は～だ」といった内容の話をするように指示をした．その後，実験参加者に対して 4 種類のモードから 2 種類のモードを抽出し，2 種類のモードに対しそれぞれ 1 分間対話させ，一対比較させた．一対比較は実験参加者 1 人につき 2 通り行った．1 モード終わる毎に，表 3.2 の 8 項目について 7 段階（中立 0）で官能評価の聞き取りを行った．同時に各モードについて感想の聞き取りを行った．実験参加者は，要支援または要介護認定を受けた方を対象としたデイサービス利用中の 74～97 歳の高齢者 31 名であり，性別は男性 7 名，女性 24 名，年代別は 70 代 2 名，80 代 18 名，90 代 11 名であった．実験の様子を図 3.7 に示す．なお，本実験は岡山県立大学倫理委員会の承認を得た上で，事前に実験参加者に目的・方法を口頭と文章にて説明し，同意を得て実施した．録画・録音されたデータは研究者以外の目に触れないこと，匿名性の保護，研究への参加を拒否した場合も不利益を受けないこと，得られたデータを研究以外の目的で使用しないこと等を保障した．

Table 3.1 システム評価実験の条件

モード	聞き手の種類	画像うなずきの有無
Aモード	バーチャルエージェント	無
Bモード	バーチャルエージェント	有
Cモード	ロボット	無
Dモード	ロボット	有

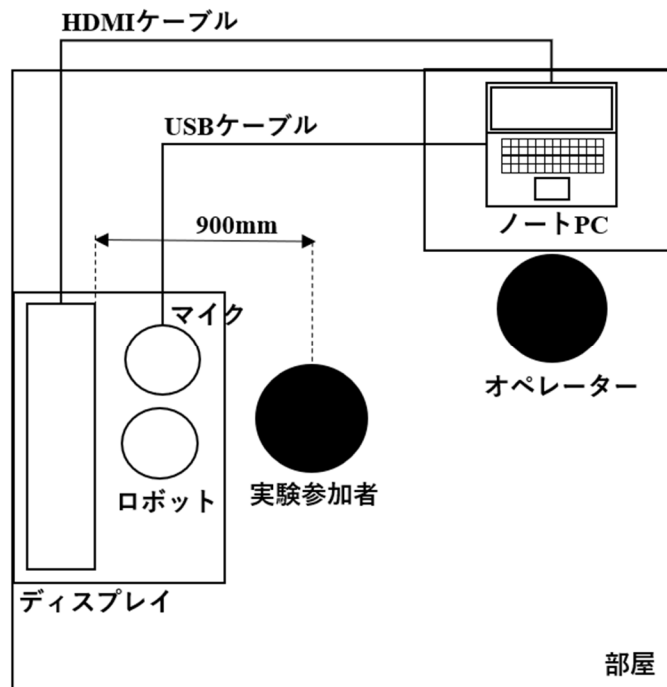


Fig. 3.6 実験システムの概略

Table 3.2 アンケート項目

No.	項目
①	楽しみを感じましたか
②	親しみを感じましたか
③	システムをまた利用したいですか
④	話しやすかったですか
⑤	1分より短く感じましたか
⑥	話が伝わったと感じましたか
⑦	話そうとする気持ちが高まりましたか
⑧	話題が頭に浮かびましたか

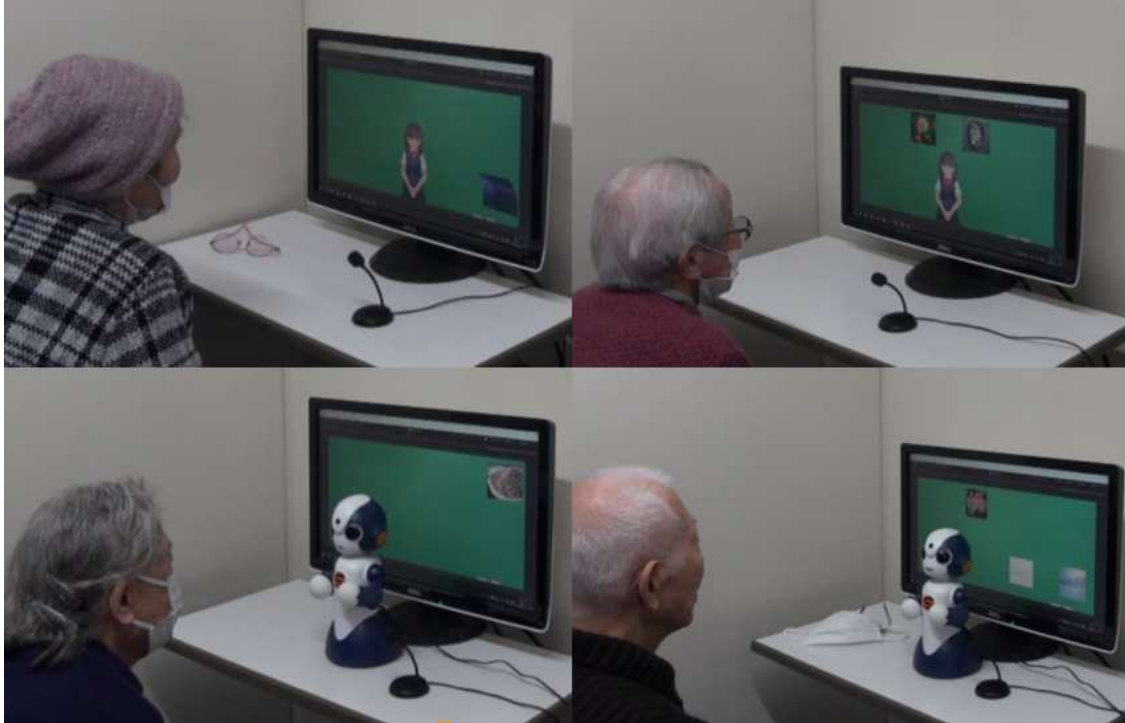


Fig. 3.7 実験の様子

3.3.2 実験結果

4種類のモードに対する一対比較の結果を表 3.3 に示す. 表中の数字は, 各列に対し各行の表示をより好ましいと答えた実験参加者の数を表している. この一対比較の結果に対して, Bradley-Terry モデル ($P_{ij} = \pi_i / (\pi_i + \pi_j)$, $\sum \pi_i = \text{const.} (= 100)$, $\pi_i: i$ の強さ, $P_{ij}: i$ が j に勝つ確率) を想定し, 各モードの強さ π を最尤推定した結果を図 3.8 に示す. これにより, 聞き手に人型コミュニケーションロボットを使用し, 画像のうなずきが有る D モードが最も高く評価されており, 続いて C モードが評価され, B モード, A モードの順に評価された.

Table 3.3 一対比較の結果

	A	B	C	D	合計
A		6	4	4	14
B	6		4	4	14
C	7	4		8	15
D	4	7	8		19

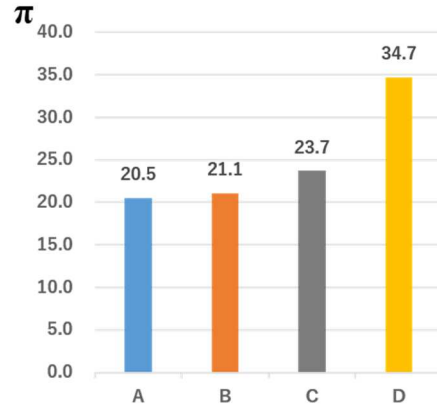
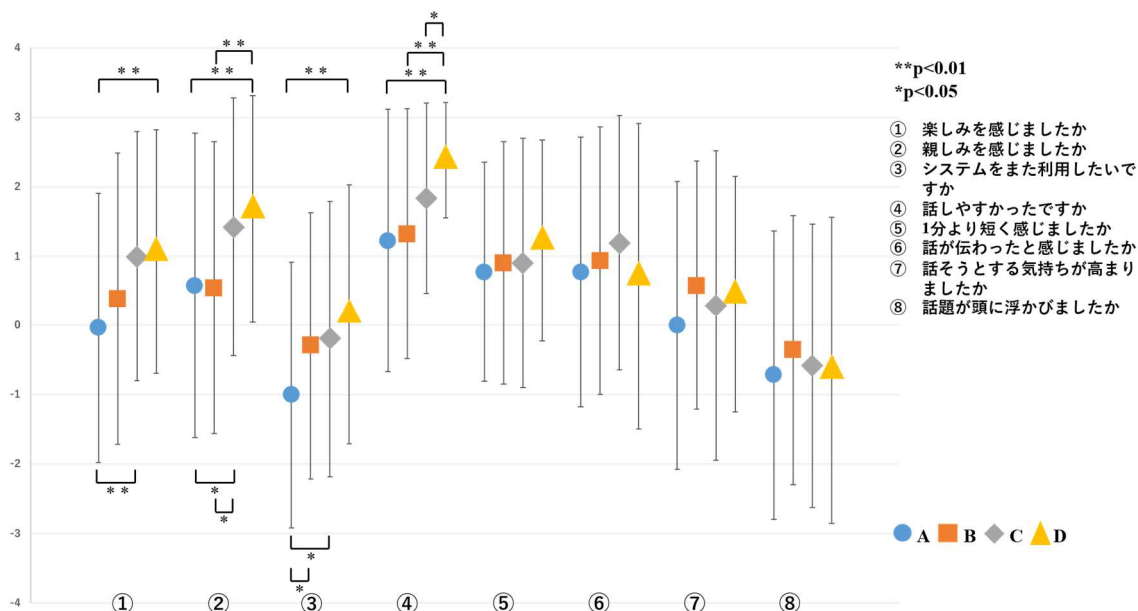


Fig. 3.8 Bradley-Terry モデルに基づく π の結果

次に、7段階評価結果を図 3.9 に示す。図には各モードの平均値とその標準偏差を示している。図より「① 楽しさを感じましたか」、「② 親しみを感じましたか」、「③ システムをまた利用したいですか」、「④ 話しやすかったですか」、「⑤ 1分より短く感じましたか」の 5 項目に一対比較実験と同様の評価傾向があることがわかる。Friedman の分散分析および Wilcoxon の符号順位検定を用いた多重比較を行った。その結果、A モードと B モードの間で、「③ システムをまた利用したいですか」の項目で有意水準 5%、A モードと C モードの間で、「② 親しみを感じましたか」、「③ システムをまた利用したいですか」の項目で有意水準 5%、「① 楽しさを感じましたか」の項目で有意水準 1%、A モードと D モードの間で、「① 楽しさを感じましたか」、「② 親しみを感じましたか」、「③ システムをまた利用したいですか」、「④ 話しやすかったですか」の項目で有意水準 1%、B モードと C モードの間で、「② 親しみを感じましたか」の項目で有意水準 5%、B モードと D モードの間で、「② 親しみを感じましたか」、「④ 話しやすかったですか」の項目で有意水準 1%、C モードと D モードの間で、「④ 話しやすかったですか」の項目で有意水準 5%で有意差が認められた。「⑤ 1分より短く感じましたか」、「⑥ 話が伝わったと感じましたか」、「⑦ 話そうとする気持ちが高まりましたか」および「⑧

話題が頭に浮かびましたか」の項目ではどのモード間においても有意差が認められなかった。



	A			B		C
	B	C	D	C	D	D
①	-	1%	1%	-	-	-
②	-	5%	1%	5%	1%	-
③	5%	5%	1%	-	-	-
④	-	-	1%	-	1%	5%
⑤	-	-	-	-	-	-
⑥	-	-	-	-	-	-
⑦	-	-	-	-	-	-
⑧	-	-	-	-	-	-

Fig. 3.9 アンケートによる官能評価の結果

また、7段階評価結果を二要因分散分析した結果を表 3.4 に示す。聞き手バーチャルエージェントか聞き手ロボットのどちらを使用するかといった聞き手の種類に関する群とうなずきの有無に関する画像群に分類したものを二要因分散分析した結果、聞き手の種類の群で「① 楽しさを感じましたか」の項目で有意水準 5%、「② 親しみを感じましたか」、「④ 話しやすかったですか」の項目で有意水準 1%で有意差が

認められた。また、画像うなずきの群で有意差は認められなかった。二要因分散分析の結果、すべての項目で交互作用が認められず、因子：聞き手の種類は「① 楽しさを感じましたか」、「② 親しみを感じましたか」、「④ 話しやすかったですか」の項目で主効果が認められた。因子：画像うなずきでは主効果が認められなかった。このことより、聞き手の種類の効果により、使用者が話しやすくなる発話促進効果があることを示した。

Table 3.4 二要因分散分析の結果

項目\因子	聞き手の種類	画像うなずき	聞き手の種類 &画像うなずき
①	0.015*	0.488	0.611
②	0.006 **	0.753	0.686
③	0.076	0.133	0.615
④	0.003 **	0.251	0.421
⑤	0.455	0.455	0.749
⑥	0.790	0.657	0.375
⑦	0.822	0.303	0.560
⑧	0.832	0.702	0.580

*p<0.05 **p<0.01

実験中の画像提示数は A モードが 51 枚、B モードが 49 枚、C モードが 33 枚、D モードが 36 枚であった。聞き手バーチャルエージェントを使用した AB モードの合計提示数は 100 枚あり、その内 39 枚に対して使用者が反応を示した。聞き手ロボットを使用した CD モードの合計提示数は 69 枚あり、その内 43 枚に対して使用者が反応を示した。使用者が反応を示した画像数合計 82 枚の内、15 枚に対して画像を凝視して話が止まることがあった。これは小さくて視認できないことや話したもののイメージと異なる画像が出てきたことによる思考が原因と考える。正確に認識できたかイメージと異なる画像であるが容認された画像数 67 枚の内、31 枚に対して使用者が画像の話題に触れる行動が見られた。聞き手バーチャルエージェントの画像誘導率は

39.0%，聞き手ロボットは 62.3%となり聞き手ロボットの方が画像への誘導率が高い結果を示した。

実験中の様子観察，感想の聞き取り調査の結果を表 3.5 に示す。様子観察と聞き取り調査の中で「画像によく反応していた」「話していたら絵が出て，もっと話したかった」といった画像誘導に関するポジティブな要因や「画像で話題変更が起こった」「しゃべったことが出ると良かった」といった画像提示に関するポジティブな要因，「聞いてくれてうれしい」「言いたいことが言えるのが良い」「ずっと聞いてくれるから良かった」といった傾聴に関するポジティブな要因，「聞き手に多くの質問をしていた」「うなずきがあるのでリズムで話せた」「何を話しても良いと感じた」といった発話に関するポジティブな要因等が見られた。しかし一方で，様子観察と聞き取り調査共に画像提示に関して「画像が出ることで話がとまった」「画像が小さくて見えていないまたは画像提示に全く気が付いていない」「絵が小さい」「話した絵が出ていない」「しゃべったものと違う画像が出ると伝わってないと感じた」といったネガティブな要因が見られ，画像のサイズによる見えにくさやイメージと異なる提示画像を容認させる方法など，高齢者特性に合わせた提示方法や容認方法が必要であると考える。また，「横を向くのが怒ったように感じた」といったネガティブな要因に対して，「画像につられて話が出た」といったポジティブな要因も見られた。これは画像への誘導の意図が上手く伝わっていないといった，実験参加者毎の提示画像への誘導に対する捉え方の違いと考える。さらに，「聞き手に多くの質問をしていた」「話してくれるのを待つ」等の様子や「声が聴きたかった」「何か話してほしい」「対話欲しい」「独り言，返答がない」等の意見が多くあり，ノンバーバルな反応だけでなく声によるあいづちや対話といったより人間とコミュニケーションをしているような反応を求めていると考える。

Table 3.5 実験の観察と感想の聞き取り

<p>ポジティブな様子</p> <ul style="list-style-type: none"> ● 画像によく反応していた ● 子供に話しかけるように楽しんで使っていた ● 画像で話題変更が起こった ● 聞き手キャラクタに多くの質問をしていた
<p>ネガティブな様子</p> <ul style="list-style-type: none"> ● ロボットが横を向いた時話が途切れた ● 画像が出ることで話がとまった ● 画像が小さくて見えていない様子 ● 画像提示に全く気が付いていない様子
<p>ポジティブなコメント</p> <ul style="list-style-type: none"> ● 話していたら絵が出て、もっと話したかった ● 画像につられて話が出た ● うなずいてくれたので理解してくれたと感じた ● ずっと聞いてくれるから良かった ● 昔のことを思い出して楽しかった ● しゃべったことが出ると良かった ● 聞いてくれてうれしい ● うなずきがあるのでリズムで話せた ● sotaは実物だから良い ● 反応が良かった、新しいものは楽しい ● 言いたいことが言えるのが良い ● 悪い事も気にしないで話せるから良い ● 楽しい思い出が出てきた ● 何を話しても良いと感じた ● 話し相手がないときには話したい
<p>ネガティブなコメント</p> <ul style="list-style-type: none"> ● 横を向くのが怒ったように感じた ● しゃべれるようになったら良い ● しゃべったものと違う画像が出ると伝わってないと感じた ● 絵が小さい ● もっと動きがほしい ● 声のあいづちがほしい ● 一人では話にくい、何か話してほしい ● 人間でないから冷たい感じで人形と話している感じ ● 若いから話しが合わないと感じた ● 対話がほしい、独り言のようだ ● 返答がない

3.4 結言

本章では、聞き手バーチャルエージェントより一体感や実在感を対話者に実感させることができる人型コミュニケーションロボットを聞き手とすることで、現実との親和性を高めることによって高齢者の発話を引き出していく傾聴システムを開発した。

本システムを用いて高齢者介護施設の高齢者を対象に評価実験を行い、人型コミュニケーションロボットを用いることで実在感のある傾聴を実現しつつ、提示画像に注視誘導し画像誘導率を高めるとともに官能評価によりシステムの有効性を確認した。

聞き手の種類の効果により、楽しさや親しみを与えることや使用者が話しやすくなる等の発話促進効果があることを示した。さらに、聞き手バーチャルエージェントの画像誘導率 39.0%に対して、聞き手ロボットは 62.3%となり、聞き手ロボットの方が画像への誘導率が高い結果が示され、画像からのインタラクション効果を得やすくなることを示した。本実験では 2 章では有意差が認められなかったアンケート項目「③ システムをまた利用したいですか」において、有意差が認められた。聞き手ロボットが現実で大きな動きで目を引くことから、聞き手がロボットではなく画像のうなずきがない A モードでは他と比べてインタラクションが少ないため物足りなさを感じ、他のモードに対して差が生じたと考える。また同じように、「⑥ 話が伝わったと感じましたか」、「⑦ 話そうとする気持ちが高まりましたか」の項目において、オンライン検索方式を採用したことからイメージと少しずれた画像が提示されることがあるため、提示画像が発話単語と合致していると認知できなかったことにより、聞き手ロボットであっても聞き手バーチャルエージェントであっても効果が認められなかったと考える。また、視力が低下した高齢者にとっては画像が小さく認知できなかったことも考えられる。今回の実験では、提示画像はオンラインから取得しているが、実験時の話題は食べ物に関するものに限定している。話題を限定しない場合、さらに多くの多岐にわたる画像が提示されるため、提示画像を認知しやすくする仕組みが必要である。

2 章で開発した傾聴システムにおいては、介護従事者を対象とした評価実験で、高齢者介護施設の現場で使えそうであるとの感想を得ており、今後の高齢者の新たな話し相手となりうる可能性を示したが、今回の実験では高齢者特性に合わせた画像提示やコミュニケーションが必要であることが確認された。聞き手ロボットによる振り向き動作だけでなく、使用者が話した内容に基づく音声によるフィードバックを用いることで、画像が提示されたことを強調させることや聞き手ロボットが話の内容を聞き取れていることを表現する方法を検討し、本システムの発展を目指す。

参考文献

- [3.1] 山崎裕之, 石井裕, 渡辺富夫, 高齢者の発話促進を目的とした発話単語のうなずき画像を用いた傾聴システム, 日本機械学会論文集, Vol. 88, No. 913, Paper No. 22-00046, pp.1-13(2022).

第4章 ロボットによるバックトラッキング機能

4.1 緒言

3章では、音声認識にて得られた発話内の単語に紐づく画像をオンライン上で検索し画面上に提示し、同時に人型コミュニケーションロボットが、振り向き動作により提示された画像に注視誘導を行う傾聴システムの開発・評価を行い、有効性を示した。しかし、聞き手ロボットによる振り向き動作だけでなく、使用者が話した内容に基づく音声によるフィードバックを行うこと、画像が提示されたことを強調させることや聞き手ロボットが話の内容を聞き取れていることを表現する方法の検討が必要となった。

心理学の分野では、コミュニケーション内容を相手にそのままフィードバックするバックトラッキングとよばれる技術がある。聞き手の価値判断や評価を脇に置いて、話し手の表明していることをそのまま受け取って返していくというものである[4.1]。繰り返されることによって、話し手も聴いてもらえたという安心やそのように自分は話しているという確認ができる[4.2]。バックトラッキングを行うことで、自分のことを聞いてもらっている、受け入れてもらっているという好感や安心感、信頼感を相手に与えることができる。

本章では、使用者の発話内容を音声認識し、発話内の単語に紐づく画像を画面上に提示すると同時に聞き手ロボットが使用者の発話単語を音声によるバックトラッキングを行うことにより、発話を促進させることを目的とする傾聴システムの開発を行う。これにより、発話単語に応じて写真や絵画の画像が次々と提示されると同時に聞き手ロボットの音声によるバックトラッキングがなされ、話題を提供しながら積極的に関与して何でも聞いてくれているという安心感を与えることで発話を促進し、高齢者介護施設での新たな対話相手となることが期待される。なお、使用者が話した内容に対してイメージと異なる画像が提示されてしまった場合、イメージと離れすぎてい

る場合やそもそも画像が小さく認知できず、使用者が聞いて理解していないと感じてしまうという課題にも対応した。

4.2 バックトラッキング機能

バックトラッキングとは、日本語では「オウム返し」といわれ、相手の発した言葉や感情を、同じように表現することで、相手の共感を得やすく、信頼関係を構築することができる心理学のスキルである。本システムでは、検出された単語に対して、聞き手ロボットが疑問文で一定の要求を相手に求めているかのようにバックトラッキングを行い、使用者が能動的に解釈することで、使用者は聞き手ロボットが自分の話を理解しようとしているという気持ちを感じることができ、安心感を得ることができ、要求の中には、正しく聞き取れて理解できているか確認を求める理解確認要求、聞き取れなかったような印象を与えてもう一度発話を求める反復要求、理解できなかったような印象を与えて説明を求める説明要求等がある。理解確認要求を重ねることで安心感を与え、自分の話を続けやすくして発話を促進させる。また、使用者が伝わっていないと感じ、別の表現で伝えようとすることで新たなコミュニケーションが生じることが期待される。本システムの使用例として、「今日の朝食はパンを食べました」といった話に対して「パン」という単語を検出して「パン？」というようにバックトラッキングを行い、同時に画面上に画像が提示される。使用者はバックトラッキングされた単語と画像が結びつき「そうそう」といった感情と共に、自分のことを聴いてもらっている、受け入れてもらっているという好感や安心感、信頼感が生まれ、発話を促進させる。

4.3 ロボットによるバックトラッキングを付与した傾聴システムの開発

4.3.1 コンセプト

本システムのコンセプトを図 4.1 に示す。本システムでは、実空間で発話音声に基づいて聞き手ロボットがうなずきによる身体引き込み反応をするとともに、音声を用いて発話単語をバックトラッキングする傾聴システムの開発を行った。使用者が発話

を行うとその発話単語に応じた画像が画面に提示され、聞き手ロボットはその画像の位置情報をもとに振り向きなどの身体動作を用いて注視誘導を行う。提示される画像へ注視誘導を行うと同時に聞き手ロボットが発話単語を音声によるバックトラッキングすることで、使用者は自分の話を聞き流されずに受け止められている感覚が得られ、相手に話が伝わっている感覚や安心感が増大し発話が促進される。聞き手ロボットがバックトラッキングにより聞き取った情報を共有していることを示すことは、聞き手ロボットの理解の根拠となる有効なフィードバックとなる。聞き手ロボットが理解していることを確認できることは、使用者に安心感を与え、発話を促進することが期待される。

本システムでは、得られた単語に対して、聞き手ロボットが疑問形で一定の要求を相手に求めているかのようにバックトラッキングを行う。使用者はその要求を能動的に解釈することで、聞き手ロボットが自分の話を理解しようとしているという気持ちを感じることができ、安心感を得ることができる。要求の中には正しく聞き取れて理解できているか確認を求める理解確認要求や聞き取れなかったような印象を与えてもう一度発話を求める反復要求や理解できなかったような印象を与えて説明を求める説明要求等がある。理解確認要求を重ねることで聞いて理解してくれるといった安心感を与え、自分の話を続けやすくして発話を促進させる。また、使用者が伝わっていないと感じ、別の表現で伝えようとすることで新たなコミュニケーションが生じることが期待される。システムの使用風景を図 4.2 に、バックトラッキングを伴う動作遷移図を図 4.3 に示す。

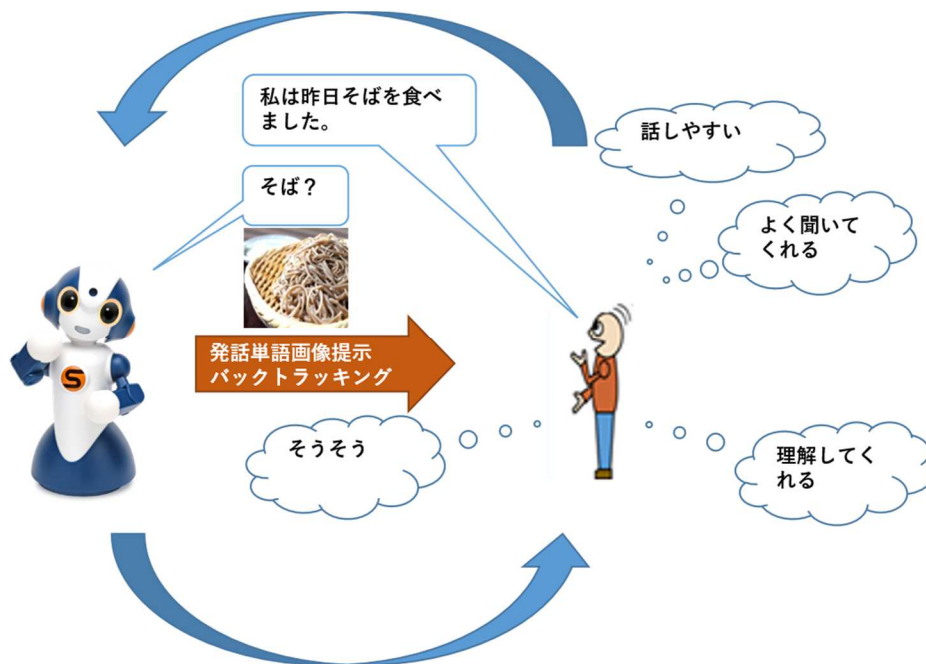


Fig. 4.1 ロボットによるバックトラッキングを付与した傾聴システムのコンセプト



Fig. 4.2 システムの使用風景.

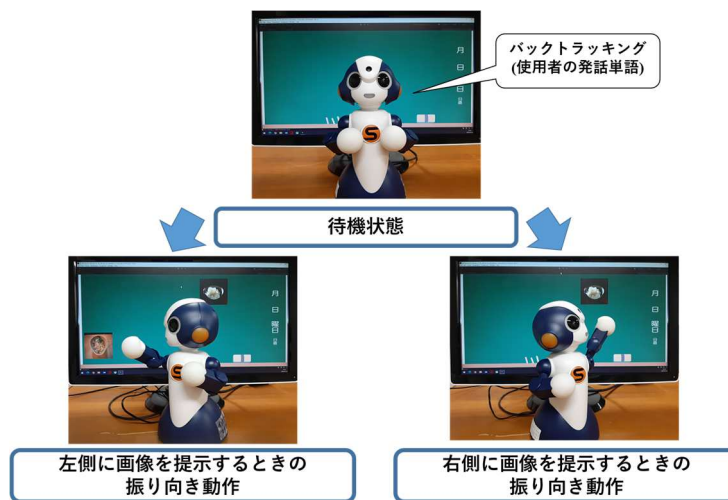


Fig. 4.3 バックトラッキングを伴う聞き手ロボットの画像への視線誘導動作の遷移

4.3.2 ハードウェア構成

本システムは、画像描画用 PC (MSI GE66 RAIDER (CPU: Intel Corei9-10980HK 2.40GHz, Memory:16GB, Graphics: NVIDIA GeForce RTX 2070 Super)) , USB マイク (SANWA SUPPLY MM-MCU08BK) , 人型コミュニケーションロボット(ヴイストーン社 Sota) から構成される。画像は画像描画用 PC から出力しディスプレイ (DELL ST2410) に提示する。解像度は 1920×1080 pixel である。画面輝度は設定可能範囲の最大値である。画面色情報において、色形式は RGB を採用し、色空間には標準ダイナミックレンジを採用した。人型コミュニケーションロボットに搭載されているインテリジェントマイク(図 4.4)により使用者が発話する音声を集音し、USB マイクによって音声の ON-OFF を集音する。

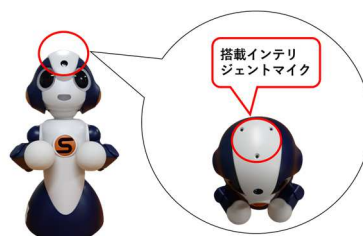


Fig. 4.4 人型コミュニケーションロボット搭載インテリジェントマイク

4.3.3 システム構成

本システムの環境構築は、画像提示処理や発話内容から単語を抽出する処理は Unity2020.2.3 を用い、一連の動作を C# で実装している。使用者の発話単語の音声認識には聞き手ロボットに搭載されているインテリジェントマイクを用いた。聞き手ロボットが搭載マイクで取得した発話音声を認識し、検知した会話の区切りでテキストデータに変換して Unity 側に情報を渡す。Unity 側にて渡された情報を基に形態素分析を行い、単語を検出する。複数の単語が検出された場合は、最後に検出された単語を採用する。画像検索 API を使用して検出単語の画像を取得し、画面に提示する。同時に抽出単語は聞き手ロボット側へテキストデータで返送され聞き手ロボットの音声によるバックトラッキングをするように実装している。聞き手ロボット側のシステム構築には Java を用いて、ヴイストン社が発行しているライブラリを使用した。聞き手ロボットのインタラクションモデルとして、音声の ON-OFF パターンに基づくうなずき反応モデルを導入している。音声認識から聞き手ロボットによるバックトラッキングが実行されるまでのデータの流れを図 4.5 に示す。

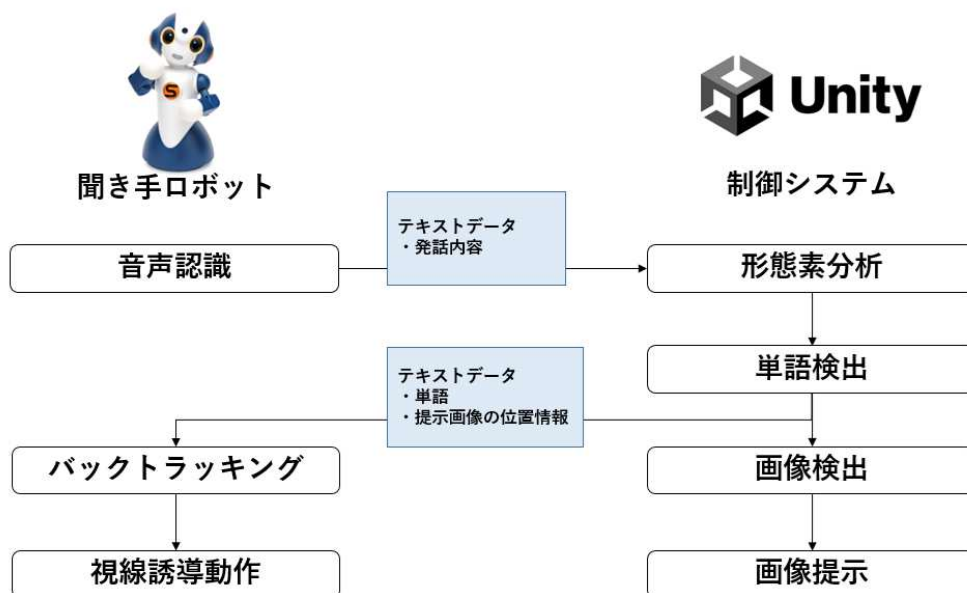


Fig. 4.5 ロボットによるバックトラッキングを付与した傾聴システムのデータフロー図

4.4 ロボットによるバックトラッキングを付与した傾聴システムの評価

4.4.1 実験方法

発話単語をバックトラッキングする機能を付加した聞き手ロボットとともに、表 4.1 に示す 3 種類のモードを用意し、評価実験を行った。A モードはバックトラッキングを行う確率 0%、B モードはバックトラッキングを行う確率 50%、C モードはバックトラッキングを行う確率 100% で実験を行った。実験システムの概略を図 4.6 に示す。すべてのモードにおいて画像提示手法として、発話単語内の単語を形態素分析によって抽出し、単語を画像 API に受け渡しオンライン上の画像を取得し提示するオンライン検索方式を用いた。提示位置はあらかじめ設定した 10 箇所ランダムに提示される。聞き手ロボットには、人型コミュニケーションロボットの Sota を採用した。使用者と対面にてコミュニケーションをとる位置関係を想定し、画面中央に配置した。また、発話単語の提示された方向に合わせて振り向くことで共同注視を行う。

実験手順は、まず実験概要を実験参加者に説明し、各モードを試行して反応の様子を確認させた。発話内容は食べ物や趣味、仕事に関する自由発話とし、例えば、「昨日の晩ご飯は～を食べた」、「好きな食べ物は～だ」、「趣味は～だ」「仕事は～をしていた」といった内容の話をしてもらうよう指示をした。その後、実験参加者に対して 3 種類のモードに対しそれぞれ 1 分間体感させ、一対比較させた。一対比較は実験参加者 1 人につき 3 通り行った。1 モード終わる毎に、表 4.2 の 8 項目について 7 段階（中立 0）で官能評価の聞き取りを行った。同時に各モードについて感想の聞き取りを行った。実験参加者は、要支援または要介護認定を受けた方を対象としたデイサービス利用中の 74～96 歳の高齢者 24 名であり、性別は男性 6 名、女性 18 名、年代別は 70 代 2 名、80 代 13 名、90 代 9 名であった。なお、本実験は岡山県立大学倫理委員会の承認を得た上で、事前に実験参加者に目的・方法を口頭と文章にて説明し、同意を得て実施した。録画・録音されたデータは研究者以外の目に触れないこと、匿名性の保護、研究への参加を拒否した場合も不利益を受けないこと、得られたデータを研究以外の目的で使用しないこと等を保障した。

Table 4.1 システム評価実験の条件

モード	バックトラッキング確率
Aモード	0%
Bモード	50%
Cモード	100%

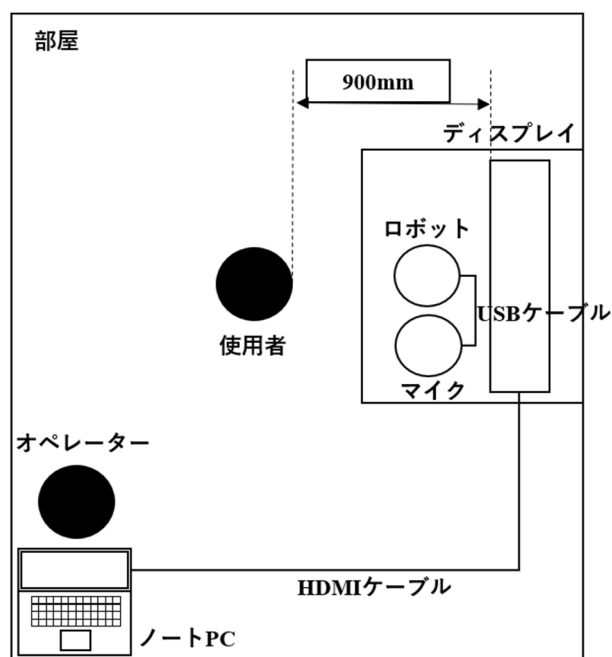


Fig. 4.6 システムの概略

Table 4.2 アンケート項目

No.	項目
①	楽しみを感じましたか
②	親しみを感じましたか
③	システムをまた利用したいですか
④	話しやすかったですか
⑤	1分より短く感じましたか
⑥	話が伝わったと感じましたか
⑦	話そうとする気持ちが高まりましたか
⑧	話題が頭に浮かびましたか
⑨	ストレスを感じましたか

4.4.2 実験結果

3種類のモードに対する一対比較の結果を表4.3に示す。表中の数字は、各列に対し各行の表示をより好ましいと答えた実験参加者の数を表している。この一対比較の結果に対して、Bradley-Terryモデル ($P_{ij} = \pi_i / (\pi_i + \pi_j)$, $\sum \pi_i = \text{const.} (= 100)$, $\pi_i: i$ の強さ, $P_{ij}: i$ が j に勝つ確率)を想定し、各モードの強さ π を最尤推定した結果を図4.7に示す。これにより、聞き手ロボットが100%バックトラッキングを行うCモードが最も高く評価されており、続いてBモードが評価され、最後にAモードという順に評価された。

Table 4.3 一対比較の結果

	A	B	C	合計
A		8	2	10
B	8		5	13
C	14	11		25

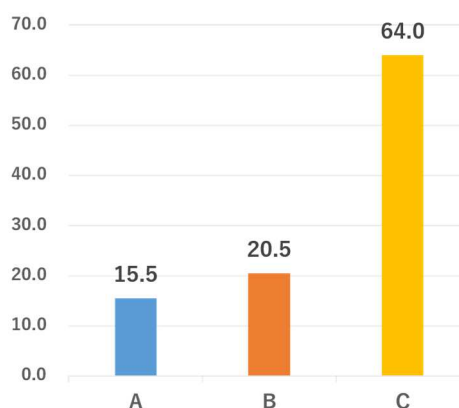
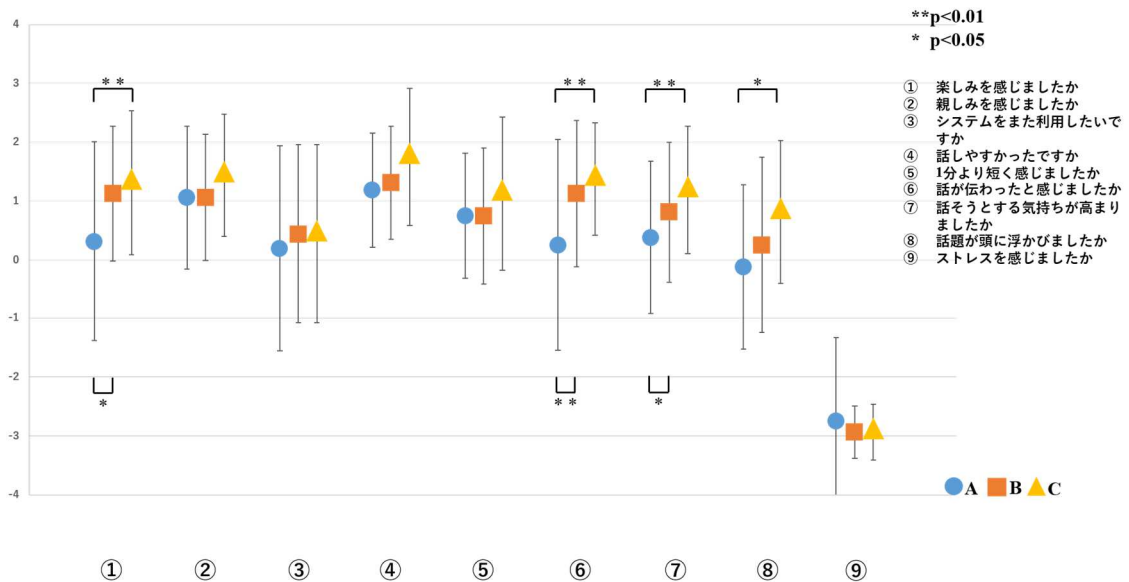


Fig. 4.7 Bradley-Terry モデルに基づく π の結果

次に、7段階評価結果を図4.8に示す。図には各モードの平均値とその標準偏差を示している。図より「① 楽しみを感じましたか」、「② 親しみを感じましたか」、「④ 話しやすかったですか」、「⑤ 1分より短く感じましたか」、「⑥ 話が伝わったと感じましたか」、「⑦ 話そうとする気持ちが高まりましたか」、「⑧ 話題が頭に浮かびましたか」の7項目に一対比較実験と同様の評価傾向があることがわかる。Friedmanの分散分析およびWilcoxonの符号順位検定を用いた多重比較を行った。その結果、AモードとBモードの間で、「⑦ 話そうとする気持ちが高まりましたか」、「⑧ 話題が頭に浮かびましたか」の項目で有意水準5%、「① 楽しみを感じましたか」、「⑥ 話が伝わったと感じましたか」の項目で有意水準1%、AモードとCモードの間で、「⑦ 話そうとする気持ちが高まりましたか」の項目で有意水準5%、

「① 楽しみを感じましたか」, 「⑥ 話が伝わったと感じましたか」, 「⑧ 話題が頭に浮かびましたか」の項目で有意水準 1%で有意差が認められた. 「② 親しみを感じましたか」, 「③ システムをまた利用したいですか」, 「④ 話しやすかったですか」, 「⑤ 1分より短く感じましたか」, 「⑨ ストレスを感じましたか」の項目ではどのモード間においても有意差が認められなかった. また, C モードにおいて100%バックトラッキングを行うことで, 使用者に煩わしさを感じさせることも考えられたが, 「⑨ ストレスを感じましたか」の項目でバックトラッキングが評価を低下させる傾向は見られなかった.



	A		B
	B	C	C
①	5%	1%	-
②	-	-	-
③	-	-	-
④	-	-	-
⑤	-	-	-
⑥	1%	1%	-
⑦	5%	1%	-
⑧	-	5%	-
⑨	-	-	-

Fig. 4.8 アンケートによる官能評価の結果

実験中の画像提示数は A モードが 54 枚, B モードが 72 枚, C モードが 56 枚であった。バックトラッキングを実施しなかった場合の合計提示数は 89 枚あり, その内 37 枚に対して使用者が反応を示した。バックトラッキングを実施した場合の合計提示数は 93 枚あり, その内 63 枚に対して使用者が反応を示した。それぞれの画像誘導率は 41.6%, 67.7%となりバックトラッキングが有る場合の方が画像への誘導率が高い結果を示した。バックトラッキングをどのように捉えたかについては, 理解確認要求が 41 回, 反復要求が 12 回, 説明要求が 0 回, 無反応, 許容が 40 回であった。

実験中の様子観察, 感想の聞き取り調査の結果を表 4.4 に示す。「バックトラッキングのタイミングで話がとまる」, 「一生懸命話しているときはうなずきだけでよい」, 「提示画像に気を取られた」といったネガティブな要因に対して, 「すぐお返事くれてよかった」, 「声を出してくれた方が分かりやすい」といったポジティブな要因も見られ, 実験参加者毎にバックトラッキングや画像提示の頻度に対する捉え方の違いが見られた。また, 「聞き手ロボットに質問をしていた」等の様子や「質問してほしかった」, 「一方的で話し合いでない」 「一方的に話すだけでは使いたくない」 「対話している感覚はない」等の意見がありノンバーバルな反応だけでなく人間とコミュニケーションを行うような反応を求めていることが確認できた。

Table 4.4 実験の観察と感想の聞き取り

<p>ポジティブな様子</p> <ul style="list-style-type: none"> ● 人間の子供のように興味を持っていた ● 「そうそう」などの肯定的な反応が見られた ● 画像の話題に変わった ● 声によく反応していた ● ロボットに質問をしていた
<p>ネガティブな様子</p> <ul style="list-style-type: none"> ● イメージと違う画像が出たときに考え込む ● 画像に全く気が付いていない、興味を持たない ● 画像がでるのを待ってしまう ● バックトラッキングのタイミングで話がとまる ● 同じ画像が何度も出て話が進まない
<p>ポジティブなコメント</p> <ul style="list-style-type: none"> ● かわいいから話したい ● 何でも言えそう ● すぐお返事くれてよかった ● 声をかけてくれるから良かった ● うなずいて伝わった気がしてうれしかった ● だっこしたい ● 通じるか不安だったがしっかりしていて安心した ● うなずきがあり親しみを感じる ● 威圧感がなく話しやすい ● 話を聞いてくれるので楽しかった ● 目が光ったりして反応があり話しやすい ● うなずきがあるから良い ● 声を出してくれた方が分かりやすい ● 話をするとよくわかってくれる ● 人間と同じような感じで話せた ● 嫌な顔せずに聞いてくれた
<p>ネガティブなコメント</p> <ul style="list-style-type: none"> ● 質問してほしかった ● 一方的で話し合いでない ● 初めてで難しかった ● 話さないといけないと感じた ● 聞く方がよい ● 人ではない感覚 ● 一方的に話すだけでは使いたくない ● 一生懸命話しているときはうなずきだけでよい ● 返事が遅い ● しゃべり返してくれないと何を考えているのかわからない ● 声が機械的 ● 画像が出てこないと伝わってないと感じた

4.5 結言

本章では、聞き手ロボットを対話エージェントとして、使用者の発話によって得られた単語を聞き手ロボットが音声でバックトラッキングする傾聴システムの開発を行った。本システムの実用性の検証を行うために評価実験として、高齢者介護施設の高齢者を対象として実験を行い、官能評価により有効性を検証した。今回の実験では、使用者の発話によって得られた単語を聞き手ロボットによるバックトラッキングを行い聞き返すことで、使用者は自分の話を聞き流されずに受け止められている感覚が得られ、相手に話が伝わっている感覚や安心感が増強され自信をもって話すことができるため発話意欲が高まり「楽しみを感じましたか」「話が伝わったと感じましたか」「話そうとする気持ちが高まりましたか」「話題が頭に浮かびましたか」の項目が高く評価された。画像誘導率はバックトラッキングが無い場合 41.6%に対して、バックトラッキングが有る場合 67.7%となりバックトラッキングが有る方が画像への誘導率が高い結果を示した。また、バックトラッキングが有ることにより、聞いてくれているといった傾聴感を与えることや通じるかわからないといった不安感を緩和させることを示した。

本実験では 3 章では有意差が認められなかったアンケート項目「⑥話が伝わったと感じましたか」において、有意差が認められた。発話から得られた単語によるバックトラッキングが理解確認要求と捉えられ、正しく聞いてもらえているといった感覚を与えたことで話が伝わっているように強く感じ、有意差が生じたと考える。また、たとえ提示された画像が使用者のイメージと異なりうまく伝わらなかったと感じ、バックトラッキングを反復要求と捉えたとしても、聞いているというフィードバックを与えることができるため、有意差が生じたと考える。

今回の実験では、提示画像はオンラインから取得し、実験時の話題は食べ物、趣味、仕事と話題の幅を広げている。話題の幅を広げたことで多くの多種にわたる画像が表示された。画像の多くは問題なく認知されたが、一部認知が難しいものが提示された場合に考えこむ様子が確認できた。

参考文献

- [4.1]坂中正義, 田村隆一, 松本剛, 岡村達也『傾聴の心理学 PCA をまなぶ』, 創元社, (2017).
- [4.2]吉田輝美, 『ポケット介護 現場で使えるコミュニケーションのコツ』, 株式会社技術評論社, (2015).

第5章 高齢者特性に合わせた統合システム

5.1 緒言

2章では、本研究のベースとなる聞き手バーチャルエージェントと発話単語のうなずき画像を用いた傾聴システムの開発を行い、3章では聞き手ロボットによる注視誘導機能、4章では聞き手ロボットによるバックトラッキング機能を用いた傾聴システムの開発を行い、有効性を検証した。なお、4章ではバックトラッキングの効果を計測するため、提示画像にうなずき反応は付与していない。

本章では、発話単語のうなずき画像、聞き手ロボットによる注視誘導機能、聞き手ロボットによるバックトラッキング機能を統合したシステムの開発を行った。使用者の発話内容を音声認識し、発話内の単語に紐づく画像を画面上に提示すると同時に、聞き手ロボットが使用者の発話単語を音声によるバックトラッキングを行い、振り向き動作により提示画像に注視誘導を行う。その提示画像と聞き手ロボットがうなずくことにより、発話促進を目的とする傾聴システムの開発を行った。これにより、話題を提供しながら積極的に関与してくれる感覚を与えることで発話が促進され、高齢者介護施設での新たな対話相手となることが期待される。これまでの実験では、画像が小さく、提示に気づかれないことや認知できないといった課題が示されてきた。これまで画面内で一度に表示できる提示画像枚数の上限を左右5枚ずつ計10枚としていたが、本システムでは、提示画像枚数を左右1枚ずつ計2枚とすることで、高齢者の視力低下や画像の認識しやすさを考慮し、高齢者の見やすさや分かりやすさに注目し変更を行った。これらの機能の統合によって、より深い共感と効果的なコミュニケーションが可能となり、ユーザーエクスペリエンスの向上に寄与すると考える。

5.2 統合システムの開発

5.2.1 ハードウェア構成

本システムは、画像描画用 PC (MSI GE66 RAIDER (CPU: Intel Corei9-10980HK 2.40GHz, Memory:16GB, Graphics: NVIDIA GeForce RTX 2070 Super)) , USB マイク①(audio-technica AT9933USB) , USB マイク②(SANWA SUPPLY MM-MCU08BK), 人型コミュニケーションロボット(ヴイストン社 Sota) から構成される。画像は画像描画用 PC から出力しディスプレイ (DELL ST2410) に提示する。解像度は 1920×1080 pixel である。画面輝度は設定可能範囲の最大値である。画面色情報において、色形式は RGB を採用し、色空間には標準ダイナミックレンジを採用した。システムの概略を図 5.1 に示す。

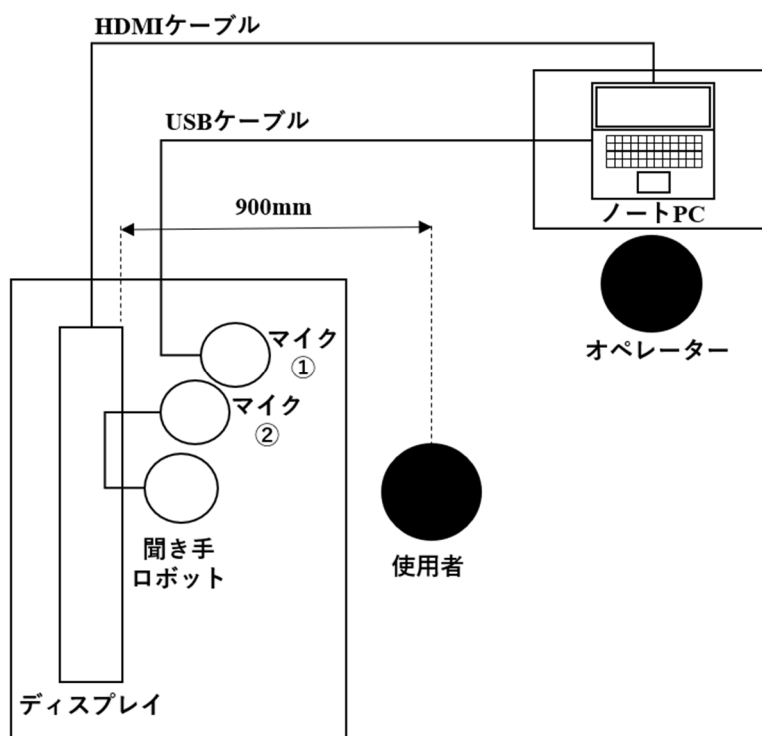


Fig. 5.1 システムの概略

5.2.2 システム構成

本システムの環境構築は、画像提示処理や画像のうなずき反応、発話内容から単語を抽出する処理等は Unity2020.2.3 を用い、一連の動作を C# で実装している。聞き手ロボットのシステム構築には統合開発環境 Eclipse にて Java を用いて開発を行い、ヴイストン社が発行しているライブラリを使用した。画像のうなずき判定を行うための集音には USB マイク①を用い、音声認識には Intel 社の RealSense SDK を用いた。聞き手ロボットのうなずき判定を行うための集音には USB マイク②を用い、音声認識にはヴイストン社が発行しているライブラリを使用した。使用者の発話単語を抽出するための集音には聞き手ロボットに搭載されているインテリジェントマイクを用い、音声認識にはヴイストン社が発行しているライブラリを使用した。聞き手ロボットの搭載マイクとマイク①、マイク②は同時に集音し、集音状況に応じて傾聴システムの一連の処理は並列で行われる。聞き手ロボットと提示画像のインタラクションモデルとして、音声の ON-OFF パターンに基づくうなずき反応モデルを導入している。

聞き手ロボットの搭載マイクを用いて、使用者の発話を音声認識し、検知した発話の区切りでテキストデータに変換して Unity 側に情報を渡す。Unity 側にて渡された情報を基に形態素分析を行い、単語を検出する。複数の単語が検出された場合は、最後に検出された単語を採用する。画像検索 API を使用して検出単語の画像を取得し、画面に提示する。同時に抽出単語は聞き手ロボット側へテキストデータで返送され聞き手ロボットの音声によるバックトラッキングをするように実装している。統合システムのアクティビティ図を図 5.2 に示す。

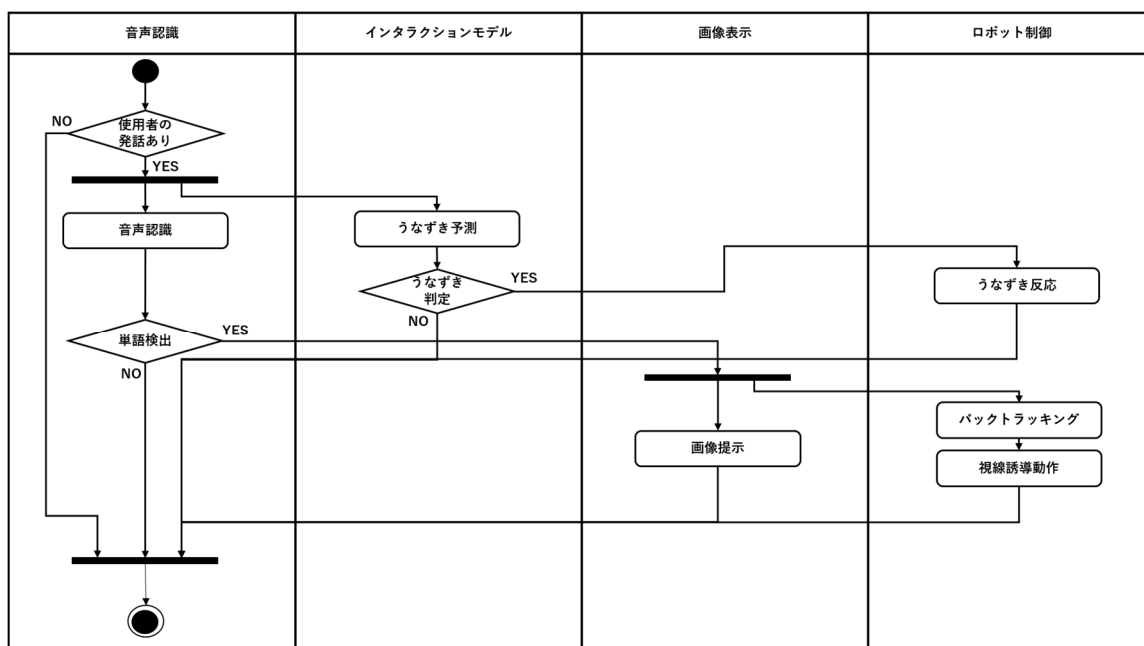


Fig. 5.2 統合システムのアクティビティ図

5.2.3 画像提示枚数

3章, 4章にて介護認定を受けた方を対象としたデイサービス利用中の高齢者に実験を行ったが, どちらの章の実験においても一定数の画像が小さく見えにくいといった様子や意見がみられた. これまで画面内で一度に表示できる提示画像枚数の上限は左右5枚ずつ計10枚であった. これは過去の画像を残すことによって話してきた内容の流れが確認でき, それを見ることで話題の発想や転換に繋がることを期待して実装したものである. しかしながら, 実際に使用したところ, 話題として取り上げられるのはほぼ最新一枚である様子が確認できた. 高齢者の視力低下や画像の認識しやすさを考慮し, 提示画像枚数を左右1枚ずつ計2枚とすることとした. 集団で身体的インタラクションを示すことでより効果があるという知見もあるが, 今回は高齢者の見やすさや分かりやすさに注目し変更を行った. 変更を加えた統合システムを図5.3に示す.



Fig. 5.3 画像提示風景 (左: 提示画像 10 枚, 右: 提示画像 2 枚)

5.3 モニタリング調査

統合システムを高齢者介護施設の一角に設置し、高齢者を対象にモニタリング調査を行った。設置風景および利用風景を図 5.4 に示す。5 組のランダムに選出した 2～3 名のデイサービス利用中の高齢者に統合システムを使用してもらい、感想の聞き取りを行った。結果を表 5.1 に示す。画像提示枚数を 10 枚から 2 枚に変更することで提示画像が見やすく気づきやすくなっており、高齢者からは「話した絵が出るのがよい」、「画像が色々でてくるから楽しい」といった画像に関する意見が多く見られた。使用中の様子観察では、提示画像が小さく認知できずに考え込む様子や提示に気づかれない様子は見られなかった。提示枚数が限定されることで統合前の傾聴システムより認知される様子が確認できた。また、2 名以上で使用したため、画像が提示されるとその提示画像を共有している様子が確認できた。一方で、「話したものと違う絵が出てきた」、「話したことがでてこない」といったネガティブな意見も見られた。これは、検出された単語に対して代表的な画像を選定し、音声のフィードバックによる支援もしているが、こちらが任意で登録したものではなくオンライン上から自動で選定したものであるため、使用者が想定しているイメージとかけ離れた画像が提示されたことによるものと考えられる。音声認識が正しくされていない様子も見られた。音声認識の仕様上、複数音源が発生する場所では正しく認識されない場合があり、2 人同時

に話しかけた場合や周りの雑音が飛び交う賑やかな場面等がそれに該当すると考える。また、「お名前はなんていうの?」、「あなたは何が好き?」等のロボットに問いかける様子が多く見られた。意見の中にも「お話がしたい」、「こちらの質問には答えてくれない」、「何を話したらいいのか分からない」等が見られた。自分のことばかり話すのではなく相手のことを知り、話のきっかけを探している様子が確認できた。



Fig. 5.4 現地設置風景

Table 5.1 モニタリング・観察と感想の聞き取り

モニタリング・観察
<ul style="list-style-type: none"> ● 提示された画像にすぐに反応しており，小さく認知できず考え込む様子は見られなかった ● 画像が提示されるたびに提示された画像を注視する動きがあり，提示を期待する様子が見られた ● 提示された画像を二人で共有して話が広がる様子が見られた ● まれに音声認識が正しくされていない様子が見られた ● ロボットに問いかける様子が見られた
使用者の感想
<ul style="list-style-type: none"> ● 画像が色々でてくるから楽しい ● うなずいてよく聞いてくれていた ● 話すと大きな絵が出て良い ● 話した絵が出るのがよい ● かわいい（ロボット） ● お話がしたい ● こちらの質問には答えてくれない ● 何を話したらいいのか分からない ● 話したものと違う絵が出てきた ● 話したことがでてこない

次に，デイサービスの運営時間中に統合システムを稼働し続け，使用状況の調査を行った．認識した日時と認識した音声をテキスト化したものを記録し，それらを用いて分析を行った．周りが賑やかな状況であったため，多量の雑音が収録された．合計18日設置した結果を集計したが，本格的な利用は確認できなかった．「おはようございます」，「こんにちは」，「やっほー」等の声かけは確認でき，中には「お名前は？」，「かわいいね」等も見られたが，いずれも座り込んで話し込むといったものではなく，通りすがりに話しかけたようなものであった．

5.4 結言

本章では，発話単語のうなずき画像，聞き手ロボットによる注視誘導機能，聞き手ロボットによるバックトラッキング機能を統合したシステムの開発を行った．また，高齢者の視力，認知力の低下を考慮し，画像の提示枚数を減らし1枚当たりの画像サイズを大きくすることで，提示画像に気づきやすく認知しやすくなるように改善を行

った。本統合システムは、高齢者に話題を提供しながら積極的に関与してくれる感覚を与えることで発話が促進され、高齢者介護施設での新たな傾聴相手となるシステムである。高齢者介護施設で使用した結果、使用者の画像への反応もよく、2人で使用した場合に画像の共有も確認することができた。これを介護現場に設置することで、周りに集まってグループで会話ができることや子供に話しかけるようにコミュニケーションができることをきっかけとして「話す」ことが促進される。また、その場にいる人と画像を共有することで話が盛り上がり楽しむことができる。通所される高齢者の一つの楽しみの方ができることを期待している。一方で、聞き手ロボットに質問するといった様子や話がしたいといった意見が多く見られ、人同士が行う対話を求めていることが確認できた。

本システムは、使用者が提示された画像を見ることで自身が盛り上がることで発話を支援することが目的であるが、話のきっかけとなる導入部や話題が途切れた接続部等はシステムの支援があるとより発話を促進することができると思う。また、現在のシステムでは、ただ設置しておくだけでは高齢者が自主的に使用することはなく、現状では、介護従事者の指示や誘導等のサポートが必要であることが確認できた。今後は、話しかけてもらえるロボットとはどういうものであるかを検討し、本システムの発展を目指す。

第6章 結論

6.1 本研究のまとめ

本研究では、使用者の発話内容を音声認識し、発話内の単語に紐づく画像を画面上に提示し、その画像と対話エージェントがうなづくことにより、発話促進を目的とする傾聴システムを開発し、その評価を行い、有効性を示した。さらに、対話エージェントの注視誘導およびバックトラッキング機能を用いたシステムを開発し、それぞれのシステムの有効性を示した。最後に、それらの機能を統合したシステムの開発を行い、高齢者介護施設の一角に設置し、モニタリング調査を行った。以下、本論文における各章の成果をまとめる。

2章では、発話内容を音声認識して、発話内の単語に紐づく画像を画面上に提示し、その提示画像がうなづくことにより身体引き込み反応を示す傾聴システムを開発した。高齢者介護施設において評価実験を行い、官能評価により有効性を検証した。検証の結果、画像提示とうなずき動作による傾聴効果により、使用者が話しやすくなる発話促進効果があることが確認できた。

3章では、聞き手バーチャルエージェントにかわり、より一体感や存在感を対話者に実感させることができる人型コミュニケーションロボットを組み込むことで、現実との親和性を高めることによって高齢者の発話を引き出していく傾聴システムの評価を行った。聞き手にロボットを用いることで実在感のある対話を実現しつつ、画像によるインタラクション効果を実空間で明示的に表現することができるのではないかという仮説を立て、実験により検証した。その結果、聞き手ロボットで提示画像に注視誘導し画像認識率を高めることで、リズム同調や親和性が増強され「楽しさ」や「親しみ」「話しやすさ」の項目が高く評価された。これらの要素が向上することで、高齢者はよりポジティブな対話体験を得ることができ、対話への参加意識が高まり発話が促進される。すなわち、人型コミュニケーションロボットを用いることで、高齢者はより実在感のある対話を経験でき、さらに人型コミュニケーションロボットが人

間のような形状や動作を持ち、同時に提示画像によるインタラクションを提供することで、対話のリアリティが向上する。これがリズム同調や親和性の増強に繋がり、対話の質を向上させたと考える。また、聞き手ロボットが高齢者の視線を提示画像に誘導することが重要であることを示した。画像認識率の向上は、高齢者が提示された画像に注目しやすくなり、それが対話の質を向上させ、これにより高齢者がより興味深く、参加しやすい対話が実現されたと言える。高齢者ケアや孤独感軽減などの分野で人型コミュニケーションロボットが有望であることが示され、特に視覚的な要素や提示画像によるインタラクションが、対話体験の豊かさや効果を向上させることを示した。聞き手ロボットを導入し、視覚的な要素を組み合わせたアプローチが高齢者との対話において効果的であることが明らかになった。

4章では、使用者の発話によって得られた単語を聞き手ロボットが音声でバックトラッキングする傾聴システムの開発を行った。バックトラッキングにより、話題を提供しながら積極的に関与して何でも聞いてくれているという安心感を与えることで発話を促進し、高齢者介護施設での新たな対話相手となることが期待されるのではないかとの仮説を立て、高齢者介護施設の高齢者を対象として実験を行い、有効性の検証を行った。実験の結果、「楽しさ」「傾聴感」「発話意欲」といった項目が高く評価されており、会話への参加意欲の高まりが見られた。これにより、聞き手ロボットにバックトラッキング機能を導入し、積極的に聞いてくれている安心感を与えるアプローチが高齢者との対話において効果的であることが明らかになった。

その結果を踏まえて、5章では、2章の傾聴システムをベースとし、3章の聞き手ロボットによる注視誘導機能、4章の聞き手ロボットによるバックトラッキング機能を統合した傾聴システムの開発を行った。その際、これまでの実験で得られた様子や意見を基に、画像枚数を減らしわかりやすくすること、画像のサイズを大きくする等の高齢者特性に合わせた改良を行った。統合システムにより、話題を提供しながら積極的に関与してくれる感覚を与えることで発話が促進され、高齢者介護施設での新たな対話相手となることが期待される。実際に改良した統合システムを高齢者施設に設置し、モニタリング調査を行った結果、提示画像が認知されないことが減少し、画像を共有することで場が盛り上がり発話を促すことが確認できた。

本研究は、画像と傾聴という 2 つのアプローチを組み合わせたシステムを提案し、高齢者の発話を促進することが目的である。2~4 章で行った検証の結果をまとめると、第一に、発話に基づく画像提示とうなずき動作による傾聴効果により、使用者が話しやすくなる発話促進効果があること、第二に、聞き手ロボットで提示画像に注視誘導し画像認識率を高めることで、リズム同調や親和性が増強され「楽しさ」「親しみ」「話しやすさ」といった効果が得られること、第三に、聞き手ロボットが音声でバックトラッキングすることで、自分の話を聞き流されずに受け止められている感覚が得られ、相手に話が伝わっている感覚や安心感が増強され「傾聴感」「発話意欲」といった効果を得られることが分かった。これらの結果は発話を促進することに寄与する。楽しさは緊張感を和らげ、楽しい雰囲気の中で行うコミュニケーションはプレッシャーが少なくなり、自分の意見や考えを気軽に発言しやすくなる。親しみは信頼感を築きやすくし、信頼できる相手に対してはより率直になることができ、自分の意見や感情を共有しやすくなる。適切な反応を示すことで話しやすい環境を提供することはコミュニケーションを円滑にし、遠慮なく発言できる雰囲気を構築する。聞き手が適切な反応によって話し手に傾聴感を与えることで、話し手は理解されていると感じ、より積極的に話そうという気持ちになる。楽しい雰囲気や親しみやすい関係、適切な傾聴により得られた発話意欲はコミュニケーションへの積極的な参加を促し、自分の考えや感情を表現したいという欲求を強める。これらの感情は相互に関連し、発話を促進する。本研究で開発した身体引き込み傾聴システムは、良好なコミュニケーション環境を整え、これらの感情を引き出すことができるものである。IT 技術を活用した画像と傾聴という 2 つのアプローチを組み合わせた手法は高齢者の発話を引き出す技術として新たな可能性を開くものと期待できる。人手不足が深刻な介護現場において、高齢者の健康を維持する「話す」機会を提供することに大きく貢献するものである。

6.2 今後の展望

本研究の目的である高齢者の発話促進に関して、本システムの発展を目指す。高齢者にとって懐かしい歌謡曲とそれに対応する画像を用意しておき、高齢者介護施設のレクリエーションで使用するといった応用も考えられる。認知症患者と介護者とのコミュニケーション支援や集団の中で人と人の会話を支援するファシリテーターのような利用方法も検討していきたい。また、高齢者が自主的に使用したくなるようなシステム、話しかけたくなるようなロボットのふるまいも検討し、システムを発展させていきたい。さらに、今後はノンバーバルな反応だけでなくより人間とコミュニケーションをしているような反応を求める要望に対し、対話型 AI を組み込むことで対話型のシステムも検討し、高齢者に話す楽しみを提供していきたい。

介護者の人手不足などの要因により、高齢者の「話す」機会が失われていく一方で現在の状況下で将来的に本システムが社会に貢献できることを期待している。

謝辞

本研究は、著者が岡山県立大学大学院情報系工学研究科に社会人として在籍していた6年間にわたり、岡山県立大学情報工学部教授、石井裕博士、特任教授・名誉教授、渡辺富夫博士のご指導の下進めてきたものです。本研究を行う上で、石井裕博士、渡辺富夫博士には、多くの時間を費やして貴重な御意見及び終始熱意に溢れた御指導、御鞭撻を受け賜りました。改めてここに深く感謝の意を表します。

岡山県立大学情報工学部教授、佐藤洋一郎博士、伊藤照明博士、春木直人博士には、本論文に関して有益な御助言を頂きました。ここに深く感謝の意を表します。

本学での研究活動に対して、深い御理解と御配慮を賜りました両備ホールディングス株式会社 小嶋光信会長、松田久副会長、松田敏之社長、および株式会社両備ヘルシーケア 小山嘉紀専務に深く感謝申し上げます。

両備ホールディングス株式会社 富田隆一郎さん、株式会社両備ヘルシーケア 藤野猛士さんには、本研究に関する論文執筆に関する有益な御助言を頂きました。ここに深く感謝の意を表します。

岡山県立大学大学院情報系工学研究科の長谷拓実さんには、本研究の一部を御協力頂きました。深く感謝の意を表します。

岡山県立大学大学院情報系工学研究科を御終了された窪田正勝さんには、本研究の一部を御協力頂きました。深く感謝の意を表します。

日頃から有意義な御助言をいただきました岡山県立大学情報工学部ヒューマンインタフェース研究室の皆様は厚く御礼申し上げます。

各種実験に御協力頂いた高齢者介護施設の介護従事者の方々及びデイサービス利用者の方々に感謝の意を表します。

最後に、著者を温かく見守り、研究に没頭する生活を可能にしてくれた家族に感謝します。

本論文に関する研究業績

原著論文

- [1] 山崎裕之, 石井裕, 渡辺富夫, 高齢者の発話促進を目的とした発話単語のうなずき画像を用いた傾聴システム, 日本機械学会論文集, Vol. 88, No. 913, Paper No. 22-00046, pp.1-13 (2022).

口頭発表

- [1] 山崎裕之, 窪田正勝, 石井裕, 渡辺富夫, 発話単語の身体引き込み画像を用いた発話促進システム, 日本機械学会第 30 回設計工学・システム部門講演会講演論文集, No. 20-14, pp.1-8 (2020).
- [2] 山崎裕之, 石井裕, 渡辺富夫, 発話単語のうなずき画像へのロボットの注視誘導による傾聴システムの開発, 日本機械学会第 32 回設計工学・システム部門講演会講演論文集, No.1407, pp.1-5 (2022).
- [3] 山崎裕之, 長谷拓実, 石井裕, 渡辺富夫, 発話単語のうなずき画像へのロボットの注視誘導による傾聴システムの評価, 日本機械学会第 33 回設計工学・システム部門講演会講演論文集, No.1504, pp.1-8 (2023)
- [4] 山崎裕之, 長谷拓実, 村上和輝, 黒川智司, 石井裕, 渡辺富夫, ロボットによるバックトラッキングを伴う発話単語画像を用いた傾聴システムの開発, ヒューマンインタフェース学会研究報告集, Vol.25, No.7, pp.1-4 (2023)