Another Meを実現する技術群

NTTが2020年に発表したデジタルツインコンピューティングのグランドチャレンジ「Another Me」では，実在する人間と同じ知性や人格を感じられ，本人として社会の中で認知され活動できる自分の分身のような存在の実現をめざします．そのための一歩として，その人らしく動作し，その人が持つ観点に沿った質問が可能なデジタルツインを構築しました．本稿では，その主要技術である，観点別質問生成技術，身体モーション生成技術，および対話映像要約技術について詳しく解説します．

大塚 淳史　高山 千尋
二瓶 芙巳雄　石井 亮
西村 徹

NTTデジタルツインコンピューティング研究センタ

はじめに

育児や介護と仕事の両立が困難となる状況や，関心や意欲があっても社会参加できないなど，人生におけるさまざまな機会の損失が社会課題となっています．活動範囲が現実世界から仮想世界へと拡大・融合する中で，人が活躍し成長する機会を飛躍的に増すために，現実世界の制約を超越して本人として活動し，活動の結果を本人自身の経験として共有できる，デジタルのもう1人の自分である「Another Me」の実現にチャレンジしています[(1), (2)]（**図1**）．このチャレンジの技術的課題としては，「人のように思考し自律的に行動が可能なこと」「本人らしい個人性を持つこと」「Another Meが得た経験をフィードバックできること」の3つが重要であり，今回はそれらを実現するための主要技術である「観点別質問生成技術」「身体モーション生成技術」「対話映像要約技術」について詳細に説明します．

観点別質問生成技術

Another Meが自律的に行動をするためには，デジタルツイン自身が次の行動について判断や意思決定を行う必要があります．そして，判断や意思決定を行うためには，判断材料となる情報を収集する手段が必須となります．私たちは，情報を収集する手段として「質問」に着目し，観点別質問生成技術を開発しました．観点別質問生成技術では，資料や会話のテキストを入力すると，入力テキストから想起される質問を自動的に生成することができます．質問を生成し，その返答を手に入れることで，デジタルツインは不足している情報を自律的に収集できるようになります．

観点別質問生成技術は，従来の質問生成技術と異なる部分が2つあります．1番目は生成する質問の内容を「観点」で制御できるという点です．質問は人の価値観や立場が大きく反映されます．例えば，社内稟議を考えたとき，営業部では価格やコストの質問が多くなり，法務部の審査では，法令順守の観点からの質問が多数を占めることが想定されます．観点別質問生成

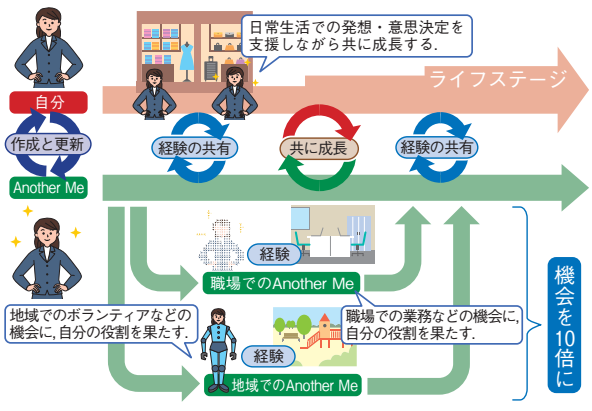


図1　Another Me

技術では，テキストと同時に観点ラベルを入力することで，入力した観点に応じた質問を生成することができます．例えば「お金」という観点ラベルを入力すると，コスト等の金額に関する質問が生成されるようになり，「法律」という観点ラベルを入力した場合には，法令やコンプライアンスに関する質問が生成されるようになります．観点別質問生成技術を組み込むデジタルツインの価値観や所属組織に応じて入力する観点ラベルを切り替えることで，デジタルツインは自身の考えや状況に対して最適な質問を行うことができ，判断に必要となる情報を収集できます．

また，観点別質問生成技術では，入力の観点とテキストの内容を理解して質問をするべきかを自動的に判断をすることができます．入力観点に関する内容がすでに記載されている場合や，入力テキストを読めば回答が分かるような場合には，質問を生成しないという機能があります．例えば，「お金」の観点ラベルを入力したとき，価格等の記載がなければ，お金に関する質問をしますが，価格やコストなどお金に関して十分な記載がある場合には質問を生成しません．これにより，デジタルツインは自身が判断を行う際に，不足している情報があるときのみ質問を行い，十分に情報が集まった際には質問をやめて次の判断の処理に移ることができます．

身体モーション生成技術

Another Meから実在する人物と同じ人格を感じられるためには，見た目はもちろんのこと，音声，発話，身体モーションがその人物らしくあることが重要であると考えられます．特に表情，顔や視線の動き，身振り手振りといった身体モーションの差異が，性格特性の差異を感じさせたり[3]，他者を識別するために大きな手掛かりとなっていること[4]を，私たちはこれまで明らかにしてきました．

このような身体モーションを，Another Meのような自律的なシステム（例えば，対話エージェント）に付与し動作させることは工学的に非常に難しい技術課題の１つです．これまで，人間らしい身体モーションや，性格特性に応じた身体モーションを発話のテキストから生成する技術[5], [6]に取り組んでいましたが，実在する特定の人物と同じようなモーションの生成は実現されていませんでした．

そこで，私たちは，日本語の発話音声情報に基づき，発話時の実在する人物らしい身体モーションを自動生成する技術を新たに開発しました．実在する人物の映像データ（音声と身体の映った画像の時系列データ）を用意するだけで，自動でその人物らしい身体モーションを生成する生成モデルを構築します．この生成モデルを利用することで，発話音声情報を入力するだけで，その人らしい発話時の動作を自動で生成することができます．技術の詳細を説明します．まず，対象となる人物の映像データに含まれる発話時の音声データから音声認識技術により，発話テキストを抽出するとともに，画像データから身体の関節点の位置を自動抽出します．次に，音声と発話テキストから身体の関節点の位置を生成可能なGAN（Generative Adversarial Networks）と呼ばれる深層学習による生成モデルを学習します．学習時に，人物の細かな癖までもとらえて幅広いモーションを生成できるモデルを構築するために，学習時にデータを上手くリサンプリングする機構に工夫があり，その人らしさや自然さといった主観評価等にて世界最高性能を保持しています（2021年10月時点）[7]．この技術をベースに，日本語音声を入力とした身体動作の生成モデルを実現しています．**図２**は，本人の入力映像，身体モーションの生成結果，入力映像の実際の正解の身体モーションの一例を示しています．

この身体モーション生成技術によって，Another MeやCGキャラクタ，ヒューマノイドロボットにおいて，特定の人物の身体モーションを自動生成させることができます．また，その他の応用先として，Web会議における本人らしいアバターの身体モーションを発話音声情報だけから簡易に自動生成できます．

今後は，身体モーション生成モデルを少量のデータで学習可能なモデルや，実在する人物の本人らしさをより追求した生成技術を構築していく予定です．

対話映像要約技術

対話映像要約技術は，録画した対話

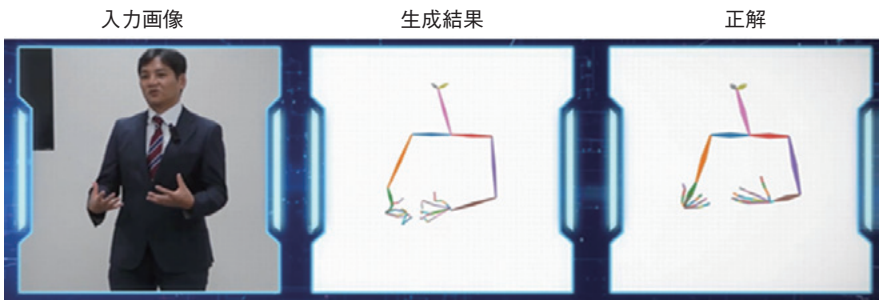


図２　本人の入力映像，身体モーションの生成結果，入力映像の実際の正解の身体モーションの一例

を実際より短い時間に要約し，内容だけでなくその場の雰囲気まで伝える映像を生成する技術です．

私たちは，人間とAnother Meが共に成長する社会を実現することをめざしています．そのためには，単に自分の代わりとしてAnother Meを利用するだけでなく，Another Meが得た経験を自分自身に効率的にフィードバックすることが求められます．また，Another Meが行ったことを，本人が「自分ごと」としてとらえられるように，内容だけではなく，その場，そのときに感じるであろう感情も伝えることも必要であると考えます．このようなAnother Meの経験を本人にフィードバックする技術の１つとして，対話を対象に対話映像要約技術の研究を進めています．

最初のステップとして，「会議の効率的な振り返りのための対話状況推定・映像要約技術」に取り組んでいます．この技術は，小型カメラやWeb会議などで撮影した会議映像を解析・再構成し，要約映像を生成します．

⑴　対話状況推定技術

対話の映像に含まれる，話者の音声や振る舞い，発言内容などのさまざまな形式の情報（マルチモーダル情報）をまとめて推定の手掛かりとして用い，発言ごとの重要度や説得力，発言の意図や意欲，参加者個人の性格特性やスキル，参加者の対話内での役割など，多様な対話の状況を推定します[(8)〜(13)]．推定では，各参加者の振る舞いの時間変化や，参加者間の動きの同期，話者音声の変化，発言内容などをまとめて学習するマルチモーダル深層学習手法や，複数の対話状況を同時に推定するマルチタスク学習手法などの機械学習の技術を利用することで，高精度な推定モデルを構築しています．

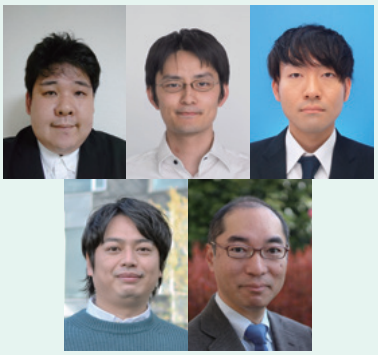
⑵　映像要約技術

前述の技術で得られた多様な対話状況の推定結果を用いて，重要な発言や，他の参加者へ問いかける発言，意見に反応する発言を抽出し，会議の映像を実時間の４分の１程度の短い映像として再構成して出力します．要約映像に含まれる参加者の表情や声のトーンから，参加者の発言の微妙なニュアンスも伝えることができます．

この技術を利用することにより，参加できなかった議論の流れや，議事録では伝わりきらない会議中の参加者の様子（意見に対する賛成・反対の態度など）や雰囲気（会議の熱量など）を短時間で効率的に把握することが可能になります．将来的には，人間どうしの対話だけでなく，人間とデジタルツインとの対話や，デジタルツインどうしの対話も要約して伝えることをめざしています．さらには，対話だけでなくデジタルツインによる行動について，より高い臨場感をもって，より効率的に人間へフィードバックする手法の研究を進めていきます．

■参考文献

⑴ https://group.ntt/jp/newsrelease/2020/11/13/201113c.html
⑵ 北原・倉橋・西村・内藤・徳永・森：“ヒトと社会のデジタル化世界を創造するデジタルツインコンピューティング構想の実現へ向けた研究開発，”NTTジャーナル，Vol. 33, No. 10, pp. 40-43, 2021.
⑶ 中野・大山・二瓶・東中・石井：“性格特性を表現するエージェントジェスチャの生成，”ヒューマンインタフェース学会論文誌，Vol. 23, No. 2, pp. 153-164, 2021.
⑷ C. Takayama, M. Goto, S. Eitoku, R. Ishii, H. Noto, S. Ozawa, and T. Nakamura：“How People Distinguish Individuals from their Movements: Toward the Realization of Personalized Agents,” HAI 2021, pp. 66-74, Nov. 2021.
⑸ R. Ishii, R. Higashinaka, K. Mitsuda, T. Katayama, M. Mizukami, J. Tomita, H. Kawabata, E. Yamaguchi, N. Adachi, and Y. Aono：“Methods of Efficiently Constructing Text-dialogue-agent System using Existing Anime Character,” Journal of Information Processing, Vol.29, pp.30-44, Jan. 2021.
⑹ R. Ishii, C. Ahuja, Y. Nakano, and L. P. Morency: “Impact of Personality on Nonverbal Behavior Generation,” Proc. of IVA 2020, No. 29, pp.1-8, 2020.
⑺ C. Ahuja, D. W. Lee, R. Ishii, and L. P. Morency: “No Gestures Left Behind: Learning Relationships between Spoken Language and Freeform Gestures,” EMNLP: Findings, pp. 1884-1895, 2020.
⑻ 二瓶・中野：“マルチモーダル情報に基づく重要発言推定モデルを搭載した議論要約ブラウザの有効性の検証，”ヒューマンインタフェース学会論文誌，Vol. 22, No. 2, pp. 137-150, 2020.
⑼ 石井・大塚・熊野・東中・青野：“話者継続・交替時における対話行為と視線行動に基づく共感スキルの推定，”情報処理学会論文誌，Vol.62, No.1, pp. 100-114. 2021.
⑽ 石井・熊野・大塚：“話者継続・交替時における参与役割に応じた視線行動に基づく共感スキルの推定，”ヒューマンインタフェース学会論文誌，Vol. 20, No. 4, pp. 447-456, 2018.
⑾ 大西・山内・大串・石井・青野・宮田：“褒める行為における頭部・顔部の振舞いの分析，”情報処理学会論文誌，Vol. 62, No. 9, pp. 1620-1628, 2021.
⑿ R. Ishii, X. Ren, M. Muszynski, and L. P. Morency: “Multimodal and Multitask Approach to Listener's Backchannel Prediction: Can Prediction of Turn-changing and Turn-management Willingness Improve Backchannel Modeling?,” Proc. of IVA 2021, pp. 131-138, 2021.
⒀ R. Ishii, X. Ren, M. Muszynski, and L. P. Morency: “Can Prediction of Turn-management Willingness Improve Turn-changing Modeling?,” Proc. of IVA 2020, No. 28, pp.1-8, 2020.



（上段左から）大塚　淳史/ 高山　千尋/ 二瓶　芙巳雄
（下段左から）石井　　亮/ 西村　　徹

デジタルツインコンピューティング研究プロジェクトでは，DTC構想の実現に向けて，企業間連携も積極的に推進しながら研究開発を進めていきます．

◆問い合わせ先
NTTデジタルツインコンピューティング研究センタ
E-mail　dtc-office-ml＠hco.ntt.co.jp