

Another Me技術による「獅童ツイン」 実現の試み

実在の人の外見・内面を再現し自律的に活動する人デジタルツインの実現をめざす「Another Me」の社会実装の第一弾として、松竹株式会社が主催する「超歌舞伎 2022 Powered by NTT」の主演である中村獅童氏のデジタルツイン「獅童ツイン」を構築し、同公演で上演しました。本稿では、取り組みの概要と、主要技術である身体モーション生成技術および任意話者音声合成技術について解説します。

ふかやま 深山	あつし 篤	いしい 石井	りょう 亮
もりかわ 森川	あきら 輝	のと 能登	はじめ 肇
えいとく 永徳	しんいちろう 真一郎	いじま 井島	ゆうすけ 勇祐
かながわ 金川	ひろき 裕紀		

NTT人間情報研究所

はじめに

デジタルツインコンピューティング (DTC) 構想のグランドチャレンジの1つである「Another Me」では、実在の人のデジタルツインが社会の中で本人に代わり活動することで、時間・空間やハンディキャップなどさまざまな制約を越えて自己実現や成長の機会を拡張することをめざしています。そのようなAnother Meを実現するための要件として「本人性」「自律性」「一体性」(図1)を設定し、その実現に向けた研究開発を進めています。

まず、Another Meが実在の人物として社会の中で活動するには、その

人の見た目や動作などの外見や、性格・価値観などの内面の再現により、その人であると認められる「本人性」が必要となります。さらに、時間や身体的・認知的なハンディキャップを越えるためには、逐一ユーザが操作したり指示したりしなくとも、その場の状況を理解し、本人と同じように判断・行動する「自律性」が必要です。これらの要件が満たされたAnother Meの活動結果から、自己実現による達成感を得たり自分自身の成長につなげたりするには、本人自らが経験したかのような実感とともに活動結果を本人に還元することで、本人のコピーであるAnother Meと本人の「一体性」を

維持する必要があります。

「獅童ツイン」実現への取り組み

3つの要件すべてが極限に満たされた存在はまさに「もう1人の自分 (Another Me)」と呼べるものになりますが、現実的には適用領域に応じてどの要件をどの程度達成すべきか決めていくこととなります。今回、Another Meの社会実装の第一弾として、本人性の実現にフォーカスし、そのための場として演劇の舞台での俳優の再現に挑みました。具体的には、伝統文化としての歌舞伎とNTTの最新テクノロジーを融合させた「超歌舞伎」に取り組みできた松竹株式会社と、超歌舞伎の主演である中村獅童氏の協力の下、中村獅童氏に代わって観客に挨拶をする「獅童ツイン」の作成に取り組みました。超歌舞伎では観客の多くが中村獅童氏や超歌舞伎のファンであるため、本人らしさ、特に外面的な本人性を高いレベルで求められる、チャレンジングな場となります。

さてここで、Another Meのよう

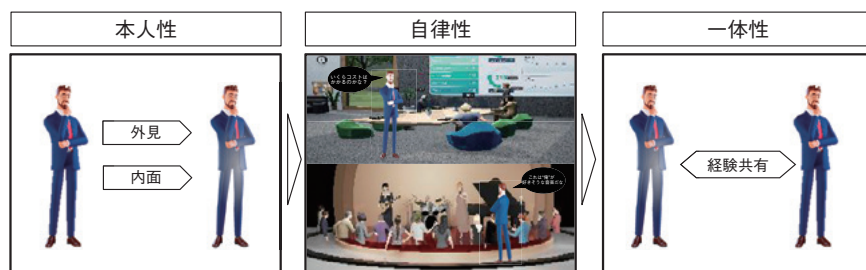


図1 Another Meの3要件

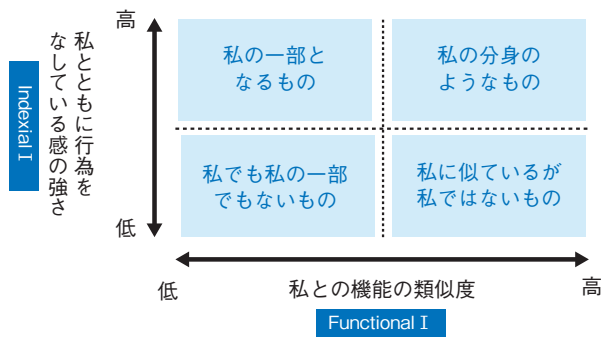


図2 「私」概念の二層構造

な、本人そのものでない存在に本人を認めるとはどういうことでしょうか。私たちは哲学の専門家との共創を通じてその要因を探究し、「機能的わたし」(Functional I)と「指標的わたし」(Indexical I)の2軸で本人性をとらえています(図2)。

「機能的わたし」とは、見た目や動作などの外面的な特徴や、スキル・能力が本人と同様であることを指しています。この観点から今回のプロジェクトでは、およそ半日間のスタジオ収録を行い、中村獅童氏の精巧な3D CGモデルを作成するとともに、本人らしい身振り・手振りと音声を生成できる機械学習モデルを構築しました。

一方で、「指標的わたし」は、その人を特徴付ける過去の経験をAnother Meから感じられるようにすることで、本人に対する指標性(「彼」「彼女」「私」など本人を指し示す意識)をAnother Meが分有し得るという考え方です。今回は超歌舞伎のファンをターゲットに、過去の超歌舞伎公演でファンにはお馴染みの観客への「煽り」を再現すべく、そのようなイメージでの身振りや発声を中村獅童氏にお願い

し、データの収録を行いました。また、伝統文化として違和感のない衣装やセリフ、舞台上の生身の共演者との関係性を感じさせる掛け合いの演出などは、松竹株式会社との綿密な検討のうえで決定しました。

以降、「獅童ツイン」を構成する技術について詳細を紹介します。

少量のデータから個人の機微な癖も再現可能な身体モーション自動生成技術

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⁽¹⁾、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⁽²⁾を、私たちはこれまで明らかにしてきました。そのため、Another Meにおいて、実在する特定の人物の身体モーションを自動的に生成させてAnother Meのモーションを制御させることが重要になりま

す。実在する特定の人物の身体モーションを自動的に生成することは工学的に非常に難しい技術課題です。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^{(3),(4)}に取り組んでいましたが、実在する特定の人物の機微な癖も模倣可能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、日本語の発話音声およびその発話テキスト情報に基づき、発話時の実在する人物の機微な癖も模倣可能な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、発話音声とその発話テキスト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、学習時に学習データを適切にリサンプリングする機構に工夫が



図3 モーション生成技術による生成されるモーションの例

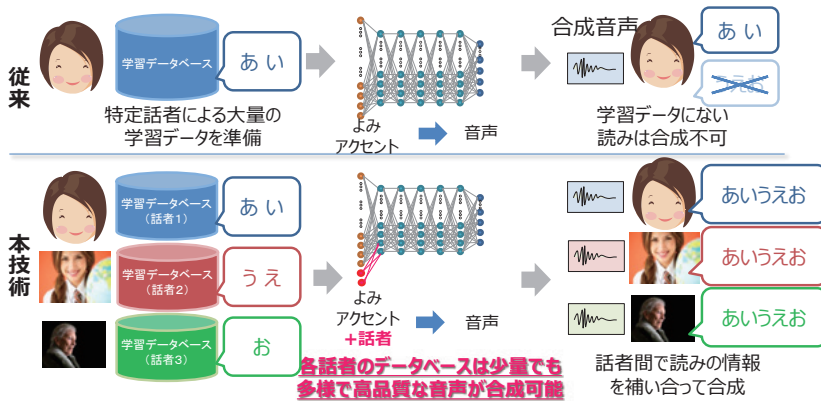


図4 多様な話者性を再現可能なDNN音声合成技術

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(2022年11月時点)⁽⁵⁾。この技術をベースに、日本語音声を入力とした身体モーションの生成モデルを実現しています。図3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。さらに現在、特定の個人の大量の映像データ(学習データ)を使用せずとも、少量のデータだけでモデルを学習可能なFew-shot learningの機構を利用した、新しい学習手法を構築しています。この手法を用いることで、中村獅童氏の口上時の10分程度の少量の映像データのみから、本人の繊細な癖までも再現可能なモーション生成モデルを構築し、

その生成結果を基に獅童ツインのモーションを制御することを実現しました。

低コストで多様な話者・口調を再現するDNN音声合成エンジン「Saxe」

本人の個性を再現するうえで、「声」は非常に重要な要素の1つとなります。音声合成技術では、再現をしたい所望の話者の音声を高精度に再現することが望まれます。しかし、所望の話者で高品質な音声合成を実現するためには、その話者が発声した大量の音声データ(例えば、波形接続型音声合成方式Cralinet⁽⁶⁾では、高品質な合成音声を生成するためには数時間~20時間程度)が必要となります。そのため所望の話者の音声合成を実現するために

は、音声収録やデータベース構築等のコストが大きな課題となっていました。この課題に対して、私たちは深層学習(Deep Neural Networks : DNN)に基づく音声合成エンジン「Saxe」を開発しています⁽⁷⁾。この音声合成エンジンでは、大量な話者が発話から構築した音声データベースと深層学習とを活用することで、所望の話者の少量の音声データからその話者での高品質な音声合成を実現しています。この方式の特徴は、大量の話者の音声データを1つのDNNでモデル化することで(図4)、読みやアクセントといった音声を生成するために必要な情報は、あらかじめ用意してある大量の話者の音声データから学習し、所望の話者の声質の特徴は、所望の話者の少量の音声データから学習します。これにより、所望の話者の音声データは少量でも高品質な音声合成を実現しています。また、本人の「声」だけでなく「演技」を再現するうえで、声色だけではなく、抑揚や発話リズム等の話し方の特徴も再現することが重要となります。しかし、演技等の非常に特殊な抑揚は発話リズムを少量の音声データから再現することは非常に難しい課題となります。これに対して、私たちは所望の話者の少量の音声データから、抑揚や発話リズムの特徴を抽出する技術を開発しています。この技術では、少量の音声を入力すると、DNNによりその音声の持つ抑揚や発話リズムの特徴を低次元のベクトルで出力します⁽⁸⁾。音声合成時には、得られた低次元のベクトルと、前述した音声合成DNNとを組み合わせることで、所望



中村獅童ツイン



澤村國矢 獅童ツイン 中村蝶紫

超歌舞伎 2022 Powered by NTT 「超歌舞伎のみかた」より © 松竹

図5 「獅童ツイン」の上演の様子

の話者の声色で、任意の抑揚や発話リズムでの合成音声の生成を実現しています。

まとめ

このような技術で実現した獅童ツインは、「超歌舞伎 2022 Powered by NTT」の演目の1つとして超歌舞伎の魅力を解説する「超歌舞伎のみかた」の中で上演され、好評を博しました(図5)。今回の取り組みをとおして、Another Me技術により商用公演に耐え得るクオリティの本人性の再現が可能であることが示されました。今後、獅童ツインの本人性やさらには自律性を成長させていくとともに、さまざまなシーンでAnother Meの社会的価値の実証に取り組んでいきます。

参考文献

- (1) 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.
- (2) 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.

- (3) 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.
- (4) 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.
- (5) 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.
- (6) 間野・水野・中嶋・宮崎・吉田：“顧客へのリアルな音声応答を実現するテキスト音声合成技術「Cralinet」,” NTT 技術ジャーナル, Vol. 18, No. 11, pp. 19-22, 2006.
- (7) 井島・小林・藪下・中村：“多様なユースケースに適用可能な音声合成エンジン「Saxe」,” NTT 技術ジャーナル, Vol. 32, No. 10, pp. 57-62, 2020.
- (8) K. Fujita, A. Ando, and Y. Ijima: “Phoneme Duration Modeling Using Speech Rhythm-Based Speaker Embeddings for Multi-Speaker Speech Synthesis,” Proc. INTERSPEECH 2021, pp. 3141-3145, Sept. 2021.



(上段左から) 深山 篤/ 石井 亮/
森川 輝/ 能登 肇

(下段左から) 永徳 真一郎/ 井島 勇祐/
金川 裕紀

Another Meはまだ生まれたての技術ですが、今後も業界を問わず積極的な連携とともに研究開発を推進し、社会に広く役に立つ技術となるよう取り組んでいきます。

◆問い合わせ先

NTT人間情報研究所
NTTデジタルツインコンピューティング研究センター
E-mail dtc-office@ntt.com