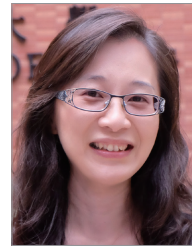


[人の感情を理解し、人に寄り添う AI]

## 5 身体動作から感情を読み取る —動作ユニット AI の構築に向けて—



藤原 健 国立中正大学 心理学系 / 東北大学 電気通信研究所  
程 苗・曾 加薫・北村喜文 東北大学 電気通信研究所

### 社会生活を支える 対人コミュニケーション

我々ヒトは「社会的動物」と言われることがあるように、他者とかかわりつながることで生活している。こうした社会生活を支えるのが対人コミュニケーションである。このことは皮肉にも、2019年から2020年にかけて始まった新型コロナウイルスの世界的な感染拡大によって改めて実感することになった。コロナ禍においても日常生活を営む上では何らかの形で対面でのコミュニケーションをとることが必要となり、その際に感染する事例が後を絶たなかったのである。また、感染拡大を防止するために積極的に導入されたオンラインでのコミュニケーション環境も、オンライン会議疲れといった新たな問題を生み出してしまった。この問題の原因の一端となるのが、対面のコミュニケーションで伝達されていた豊富な情報、特に非言語情報がオンラインでは利用しづらいことにある。

### 言語的情報と非言語的情報

対人コミュニケーションにおいては、話者たちの

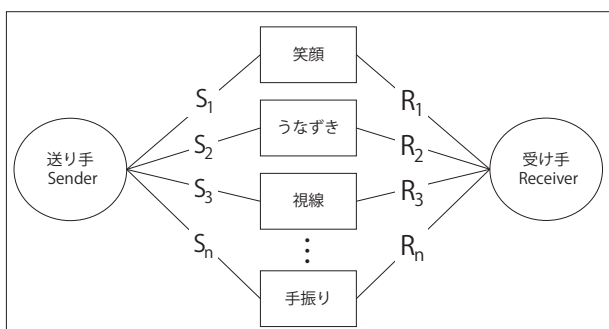
間でさまざまな情報が伝達される。言葉を介して伝えられる情報は言語的情報と位置づけられ、「明日、仙台に行きます」といったように文字に書き起こすことができることに特徴がある。論理的なメッセージを伝える際に言語的情報は欠かせない一方で、我々のコミュニケーションは言語的情報以外にも実に多くの情報をやりとりする。

言語的情報に含まれないものはすべて非言語的情報と呼ばれる。何を言うかではなくどのように言うかが非言語情報として位置づけられ、顔表情がその代表例として挙げられる。たとえば、仙台への出張を控えた人の顔に笑顔が浮かんでいれば、その人が「私は出張が楽しみです」と言葉で説明しなくてもその感情を（文字通りに）見てとることができる。反対に、その人が浮かぬ表情をしている場合には出張に対して前向きではないことをうかがい知ることができる。表情のほかにも頷きやジェスチャーといった身体動作、姿勢、近言語的特徴（発声にかかわる韻律的要素）のほかに、服装といった外見的特徴も非言語情報であり、感情という情報を伝える際には言葉の果たす役割は小さく、多くが非言語的なコミュニケーションを介して伝達されることも知られている。

## レンズモデル

非言語的コミュニケーションを理解するためのモデルとして、レンズモデル<sup>1)</sup>が一般に利用される(図-1)。このモデルでは、送り手から発信される情報が非言語行動に媒介される形で受け手に伝わることを想定する。このモデルで重要となるのが、送り手側の失敗(妥当な手がかりを表出しない、図-1の $S_1$ - $S_n$ 係数で表される)によっても受け手側の失敗(手がかりをうまく利用できない、図-1の $R_1$ - $R_n$ 係数で表される)によっても正確な読解が阻害され得ることを指摘する点である。情報が送信者の意図通りに伝達されない、あるいは受け手が誤って情報を解釈することについて、環境ノイズといった外的要因に原因を求めるのではなく、送り手・受け手両方のスキルを考慮する点で心理学的な発想をうまく取り込んでいる。また、単純な失敗に加えて、たとえば送り手が状況に沿うように非言語手がかりを意図的に操作する場合(他人をかばうために嘘をつくときなど)も考えられる。

多くの心理学研究では役者の演技を採用することで標準化された刺激を作成し、これを受け手に提示してきた。そうすることにより、送り手側の伝達失敗を極力少なくし読解の正確さを受け手側に委ねることができる。しかし、日常生活を考えた際には送り手の失敗によっても正確な理解が妨げられることがあることは念頭に置いておくべきである。送り手・受け手双方が相まってコミュニケーションが成立していることを忘れてはならない。



■ 図-1 レンズモデルを非言語的コミュニケーションに適用した例

## 感情のコミュニケーション

### 顔表情

#### 顔表情で伝える, 顔表情から読み解く

多岐におよぶ非言語情報の中でも、心理学の研究は伝統的に顔表情に注目してきた<sup>2)</sup>。顔表情と感情の関係は、古くは19世紀のDarwinに遡ることができる。そこではヒトの顔表情による感情表出は進化の過程で獲得されたものであると主張された。我々ヒトやその祖先は生存し子孫を残す過程で感情反応の恩恵を受けており、(闘争・逃走などの素早い反応など)顔表情による感情の伝達もその一端に位置すると考えたのである。具体的には、ヒトは進化の過程で感情経験と顔表情表出をつなぐ神経基盤を獲得し、これによって特定の強い感情(情動)を感じることでこれに対応する表情筋が活性化し、つまり、顔表情を介して感情を表出する機構を生得的に備えたと説明する。

そして、感情的なメッセージを読解する側にも同様の神経基盤が備わっていることから、他者の顔表情表出を見ることで感情状態を理解することができる。事実、他者の顔表情を見ると自身の表情筋も活性化し、それによって顔表情の理解が促進されることを示す研究も数多い。一般にこの顔表情による感情伝達はヒトにおいて普遍的であると考えられている。特に、喜び、悲しみ、怒り、嫌悪、恐れ、驚き、の6つはいわゆる「基本」情動として知られており、これらの顔表情は通文化的に表出され、また理解されることが多くの研究で示されてきた。

#### 顔表情を分析する

顔表情の研究では表情を司る筋肉(表情筋)を分析対象とする。ただし、顔面に存在する約40の表情筋の運動すべてを測るわけではない。各表情筋の運動を感情表出として理解する際にはアクションユニット(Action Unit)と呼ばれる分析単位に落とし込んで分析する。たとえば、喜びの感情はアクションユニット6番(口角の引き上げ)と12番(目尻の

下がり)の組合せとして表現される。現在 28 のアクションユニットが顔表情分析のために用意されている。

アクションユニットが定義されることで顔表情研究は客観性と汎用性が高まり、世界的に普及した経緯がある。筋電計を用いてアクションユニットに該当する表情筋の活動を直接記録する研究もあれば、画像ベースでアクションユニットの活動を判断する研究もあり、それぞれにアクションユニットという共通言語を用いることで顔表情研究の展開を支えてきた。

#### データ駆動型アプローチによる顔表情研究

顔表情研究の伝統的なアプローチは、人が何かしらの感情を感じた際にそれが顔表情に表れることを仮定している。反対のプロセス(顔表情を動かすことで感情状態が変わる)も表情フィードバックとして継続的に検討されてはいるものの、多くは感情が顔表情に表れることを前提としている。進化論にルーツを持つこうした理論駆動型の検証とは別に、近年ではアクションユニットを用いたデータ駆動型のアプローチによって感情理論の再構築も試みられている。

アクションユニットを用いたデータ駆動型アプローチでは、顔表情生成器がいくつかのアクションユニットを無作為に選び出し、どのタイミングでどの速さで動くかという時間パラメータに沿ってそれらを動かすことで顔表情(のような)刺激を大量に作成する。基本感情の表出のようにアクションユニットの典型的な組合せが選ばれることもあれば、感情表出という観点からはまったく異質に見える刺激も生成される。実験参加者はこれらの刺激がどのような感情表出に見えるか、その強度とともに報告するよう求められる。

既存の感情理論からアクションユニットの動きを予測し検証する理論駆動型アプローチは、科学的検証の王道である。他方、データ駆動型アプローチはどのようなアクションユニットの動きの組合せがど

のような感情表出に見えるのかという情報を(ほとんど)制約のない状態で提供してくれることから、得られる知見の柔軟さは比類がない。こうした趣きの異なるアプローチが感情理論の議論の中で有機的に融合できるのも、顔表情研究においてアクションユニットという分析単位が確立しているからであるという点はきわめて重要であろう。

## 身体動作

### 身体動作による感情表現の可能性

感情を伝えるのは顔表情だけではない。姿勢や身体動作からもヒトは感情を読み解くことができる。たとえば、肩を落とす、という表現で表されるような力が抜けて肩が垂れ下がった動作からは悲しみの感情を読み取ることができるし、全身で喜びを表現する、といった言い回しがあるように、腕を頭上に向かって大きく広げる動作からは、喜び(と興奮とが入り混じった)感情を見てとることができる。身体動作による感情表出は表情に比べて動きが大きくよりダイナミックであることが多く、誇りや恥といった一部の感情ではすでに偶然レベルの一致を超えて受け手に解読されることも知られている。

また、表情と同様に先天盲のアスリートが晴眼のアスリートと同様の身体動作で特定の感情を表現することから(競技で勝った後に両手を頭上で広げる等)、身体動作による感情表出もヒトに遺伝的に備わっていることも指摘されている。

### 現状とその問題

ただし、表情の研究に比べると身体動作による感情表出を扱った研究はあまり注目されてこなかった<sup>3)</sup>。その原因の一端として考えられるのは、身体動作の測定の難しさである。既存のコーディングシステムの中で最も知られているもの1つは The Body Action Posture coding system<sup>4)</sup>である。BAP はきわめて網羅的で、合計 141 の身体動作を対象とする一方で(e.g., 顔を前に向ける, 右手を垂直方向に上げる), 手動のコーディングシステムであることから費用対効果の

面で現実的な選択肢ではない。具体的には、訓練を積んだコーダーでさえ約 2.5 秒間の動作を解析するために 15 分かかると報告されている。BAP はごく少数のサンプルを対象にした研究では真価を発揮すると考えられるが、汎用的な利用や大規模データへの適用には向かない。

センサ等を利用して感情表出に必要となる身体動作を測定する場合、身体には稼働可能な部位が多くあり、かつそれらをどの方向に動かすのか（各関節の屈曲・伸展、内転・外転、内旋・外旋）といった運動要素との組合せを計測しなければいけないため、それだけ測定が複雑になる。ある一部の身体部位にジャイロセンサなどを取り付けて疑似的に身体動作データと捉えることもできるが、本当にその部位に感情表出のための動作が生じるのかなど、妥当性の面で疑問が残る。そのため、必然的に高精度で大掛かりなモーションキャプチャ装置で全身動作を測定することが必要となり、一般には入手が困難となる。近年、OpenPose<sup>5)</sup> といった画像ベースの姿勢推定技術が開発されて身体動作の測定自体はコストが下がりがつつあるものの、問題は測定の複雑さだけにとどまらない。

身体動作が何らかの測定値として得られたとしても、現段階では分析に落とし込む際の単位が定められていない。顔表情研究におけるアクションユニットに該当するような分析単位は身体動作の研究には

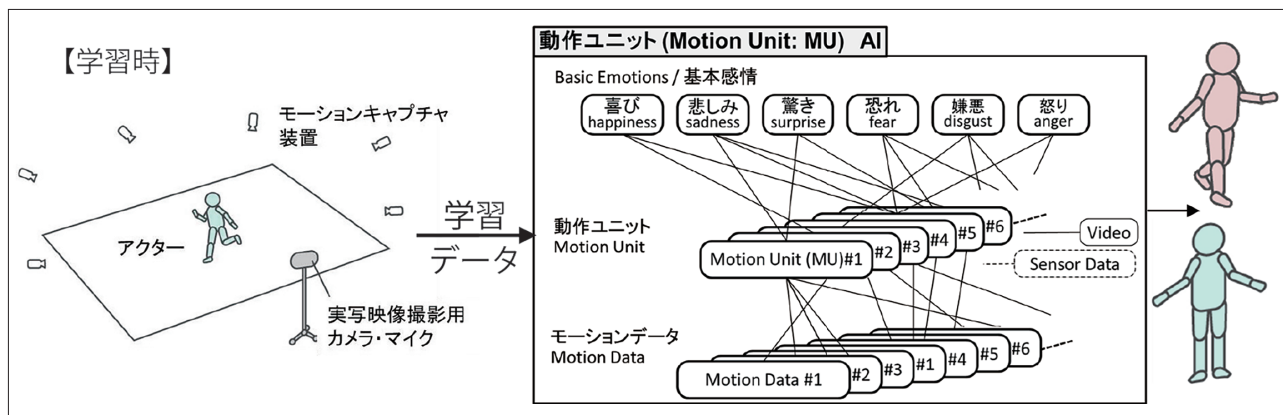
2022 年現在でも存在しないのである。研究者間に分析のための共通言語が存在しないことは、異なる環境で得られたデータ（たとえばモーションキャプチャと OpenPose）の比較や解釈を難しくし、研究の発展を大きく妨げている。

## 動作ユニットの定義に向けて

これら先行研究の問題点と感情コミュニケーションにおける身体動作の重要性を鑑みて、我々は現在、動作ユニット (Motion Unit) という分析単位を定義すべく研究を進めている (図-2)。また、将来的にはこれを自動的に判別する AI の構築も視野に入れている。動作ユニットを定義するという発想は、顔表情研究がその分析単位としてアクションユニットを定義したことで客観的な検証が可能になったことに倣っている。具体的には、感情を表出する動作を客観的・体系的に記述するために必要となる関節の運動要素とその組合せを最小限のレベルで明示・リスト化することで、世界中の研究者たちが用いることのできる共通言語を作成する。そのために、我々は現在モーションキャプチャ装置を使ってデータの取得を進めている。

## 身体動作に対するデータ駆動型アプローチ

動作ユニットを定義する際、身体動作による感情



■ 図-2 動作ユニットの定義と動作ユニット AI の構築イメージ

表出については先行研究の数が限られることから、仮説を立ててこれを検証していく従来タイプの理論駆動型アプローチでは分が悪い。代わりに我々は、モーションキャプチャから得られる大規模データを利用して、2段階のデータ駆動型アプローチを採用することとした。

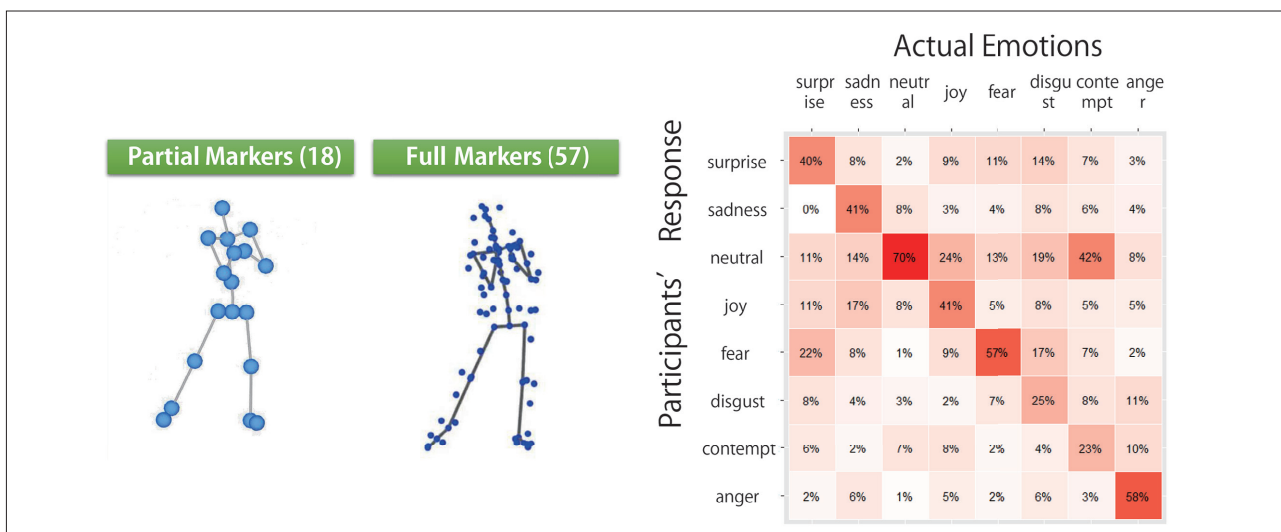
まず第1段階では、感情表出に必要な動作とそのための分析単位を抽出する。これにより、どのような感情表出にどのような身体動作が必要になるのかを見出すことができる。そしてその先には第2段階として、表情研究における顔表情生成器のように動作ユニットを操作する身体動作生成器を構築し、どのような動作ユニットの動きの組合せがどのような感情表出に見えるのかを検証していく。

動作データとしては、レンズモデルの観点と先行研究での傾向からプロの（あるいはプロに準じる）役者を日本で50名、台湾で50名雇い、それぞれに12個の感情（喜び、悲しみ、怒り、嫌悪、恐れ、驚き、軽蔑、感謝、罪悪、嫉妬、誇り、羞恥）を3つの異なるシナリオの元で3つの強度で演じ分けてもらっている（ $12 \times 3 \times 3 = 108$ ）。これにニュートラルの表出を3シナリオ分加え、1人につき111の動作データ、全体では11,100の動作データの取得を見込んでいる。

### 評価実験

レンズモデルの観点に立つと、プロの俳優から感情を表す動作データが取得されたとしても、その動作が受け手にとって正しく解読されるかどうかは保証がない。将来的に動作ユニットが定義され、AI開発へと進んでいく実用的側面を考えると、身体動作によって表出された感情は受け手に正確に解読されることが望ましい。そこで Cheng et al. ら<sup>6)</sup>は、日本で得られた動作データの妥当性を検証するために、データの中から8つの感情（喜び、悲しみ、怒り、嫌悪、恐れ、驚き、軽蔑、ニュートラル）を対象にして、3つの異なるシナリオで演じられた中程度の強度の動作データを6名分、計144の動作を2種類の Point-Light Display の形式（18点群、57点群）、つまり各関節位置を点群によって表現した身体モデルで提示した。20名の参加者に動作とともに8つの選択肢を提示して正解を強制選択するように求めた結果、正解率は約25%から70%程度になることが分かった（図-3）。

文献6)で報告されているのは暫定的な検証結果ではあるものの、選択肢が8つであることを考えるとすべての感情動作について偶然レベルよりは高い正解率が得られている（ $100/8 = 12.5\%$ ）。このことから、身体動作のみから感情を読み取ることは十分



■ 図-3 Point-Light Display 形式の身体動作表示例と 57 点群表示における各感情動作の正解率<sup>6)</sup>

に可能であること、そして動作ユニットの定義とこれを判別するAIの構築が理論的にも実用的にも有意義であることが示されたといえる。

## 動作ユニット研究の描く未来

### 感情研究の発展

モーションキャプチャによる厳密な動作データとデータ駆動型アプローチの成果として動作ユニットが定義された後は、動画ベースでも動作ユニットの活動を測定することも可能になる。OpenPoseといった姿勢推定アルゴリズムの普及に伴い、身体動作の取得は情報科学者のみならず心理学者の手にも届くようになってきている。動画から動作ユニットが抽出されることにより、従来の顔表情研究がそうであったように、身体動作による感情表出の研究がより客観的で一般的になることが見込まれる。これまであまり焦点のあたってこなかった、感情表出における身体動作の役割が明らかになる。

また、身体動作だけでなく、その他の非言語情報との関連についても理解が進むことになる。感情は、表情や音声、身体動作などの個別のチャンネルで別々に表現されるものではなく、それぞれが関連するマルチモーダルな表現として捉えられるべきであろう。一部の研究を除き、マルチモーダルな感情表出の中に身体動作の役割を積極的に見出した研究はほとんどない。動作ユニットが定義されることで感情研究の飛躍的発展が期待できるのである。

### 関連研究への波及性

顔表情研究の広範性・拡張性を見れば、動作ユニットを伴う身体動作研究がどのような分野に波及的効果をもたらすのかを想像しやすい。顔表情は感情研究の文脈だけで研究が進められているわけではなく、たとえば性格といった比較的安定した個人特性と顔表情の表出傾向には一定の関連が見られているし、嘘を見破るといった欺瞞検知の文脈でも表情に注目する研究は少なくない。また、相手を好ましく

思うか、信頼できるかといった社会的判断の研究でも顔表情が刺激として用いられる。こうした顔表情が担ってきた非言語情報の多くは、動作ユニットに置き換えて検討できる対象となり得る。

そのほか、個々人を対象とするだけでなく、対人コミュニケーションの文脈でも動作ユニットは有用である。たとえば顔表情研究の中では話者間に同一の顔表情が見られる表情模倣 (facial mimicry) という現象が知られており、良好な対人関係を予測するうえで重要な変数であると考えられている。身体動作に着目する行動模倣 (behavioral mimicry) 研究も知見の蓄積こそ多いものの、対象となる身体動作は研究者たちが恣意的に決めてきている (腕を擦る、足を揺らすなど)。動作ユニットを用いることで行動の組合せを系統的に把握できるようになることから、どのような行動が模倣されることでその後の対人関係にどのような影響が及ぼされるのかななどを包括的に理解できる可能性がある。

### 遠隔コミュニケーションへの展開

動作ユニットの活動と感情表出の関係が対応付けられ、図-2のイメージを示すような動作ユニットAIが完成した後は、キャラクターやアバタを介して特定の感情を表現するモーションをゼロから出力することが可能になる。あるいは、あるモーションデータやそれに類するセンサデータ (映像・音声や心拍などの含む) が得られた後に、これを基にして特定の感情を表現するように出力に編集を施すこともできる。動作ユニットAIは、キャラクターがただ人の動作を模した表現を出力するにとどまった従来の手法を超えて、より豊かな感情表現を体現するのに資する。これにより、制限を受けた (オンラインでの) 対人コミュニケーションの場面でも不足する情報を補って伝送するなどの手段を講じることができ、コミュニケーションを豊かにするだけでなく、対人間に生じる葛藤を解消して精神的健康を維持し、円滑な意思疎通を基盤とした豊かな社会を実現することに寄与できる。

特集  
Special Feature

また、社会実装の面でも多様な応用例が考えられる。例を図-4に示すが、たとえば、オンラインライブのイベントなどの作業工程をきわめて簡素にすることが可能になる。近年、オンラインのライブイベントは大きな商業規模を誇るようになっているが、視聴者の高い満足につながる十分なクオリティを維持するためには「舞台裏」で多くの手順が存在し、それだけ機材の数も技術者の数も必要となるのが現状である。動作ユニット AI によって推定された感情から動作アニメーションが生成できれば、リアルタイムの映像伝送を伴わずに、リアルタイム音声および制御信号のみを送信するオンラインライブが実現できるようになり作業工程の大幅な効率化が期待できる。また、最近普及しつつある VR を用いた遠隔通信であるメタバースなどでも、参加者の感情に対応するアバタの豊かな動作生成にも利用できると考えられる。

#### 参考文献

- 1) Brunswik, E. : The Conceptual Framework of Psychology, University of Chicago Press (1952).
- 2) Plusquellec, P. and Denault, V. : The 1000 Most Cited Papers on Visible Nonverbal Behavior: A Bibliometric Analysis, Journal of Nonverbal Behavior, 42(3), pp.347-377 (2018), <https://doi.org/10.1007/s10919-018-0280-9>
- 3) Witkower, Z. and Tracy, J. L. : Bodily Communication of Emotion: Evidence for Extrafacial Behavioral Expressions and Available Coding Systems, Emotion Review, 11(2), pp.184-193 (2019), <https://doi.org/10.1177/1754073917749880>

- 4) Dael, N., Mortillaro, M. and Scherer, K. R. : The Body Action and Posture Coding System (BAP): Development and Reliability, Journal of Nonverbal Behavior, 36(2), pp.97-121 (2012), <https://doi.org/10.1007/s10919-012-0130-0>
- 5) Cao, Z., Hidalgo, G., Simon, T., Wei, S. E. and Sheikh, Y. : OpenPose : Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields, IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(1), pp.172-186 (2021), <https://doi.org/10.1109/TPAMI.2019.2929257>
- 6) Cheng, M., Higashiyama, S., Fujiwara, K., Tseng, C. H. and Kitamura, Y. : E-Motion: A Database of Bodily Expression of Basic and Social Emotions, iPerception (in press).

(2022年11月7日受付)

■藤原 健 psykf@ccu.edu.tw

国立中正大学 心理学系 助理教授。博士 (人間科学)。2013年大阪大学大学院人間科学研究科博士後期課程修了。同年京都大学経営管理大学院研究員, 2014年大阪経済大学講師, 2019年カリフォルニア大学サンタバーバラ校客員研究員を経て2021年より現職。

■程 苗 cheng.miao.c3@tohoku.ac.jp

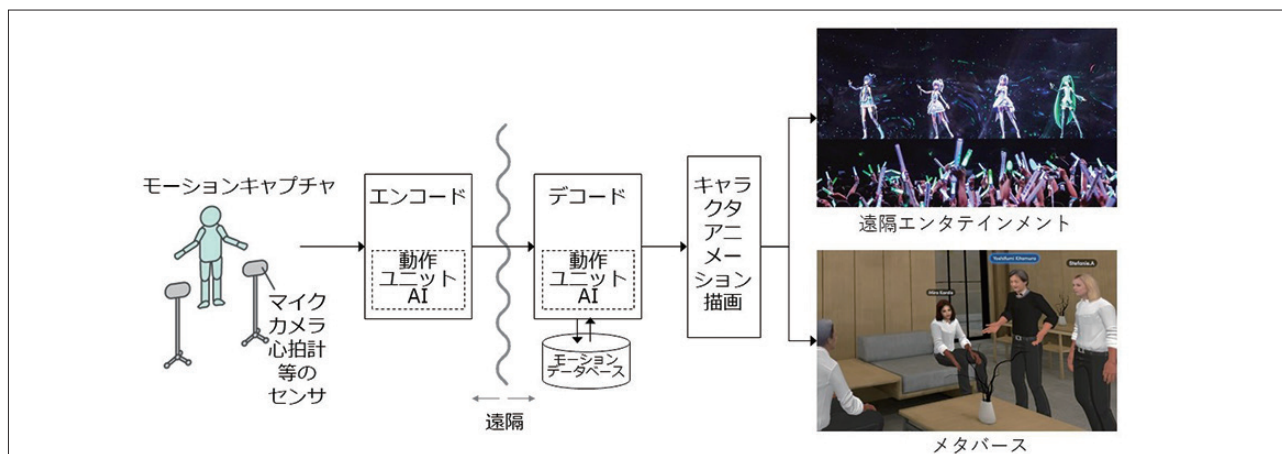
東北大学 電気通信研究所 特任助教。博士 (心理学)。2016年カリフォルニア工科大学客員研究員を経て2018年香港大学博士課程修了。同年 NTT コミュニケーション科学基礎研究所研究員。2021年より現職。

■曾 加蕙 tseng@riec.tohoku.ac.jp

東北大学 電気通信研究所 准教授。博士 (心理学)。2004年カリフォルニア大学アーバイン校博士課程修了。同年ラトガース大学研究員, 2006年国立成功大学助理教授, 2007年国立臺灣大学助理教授, 2009年香港大学助理教授を経て2016年より現職。

■北村喜文 (正会員) kitamura@riec.tohoku.ac.jp

東北大学 電気通信研究所 教授。博士 (工学)。1987年大阪大学大学院基礎工学研究科博士前期課程修了。同年キャノン (株), 1992年 ATR 通信システム研究所, 1997年大阪大学大学院工学研究科 / 情報科学研究科 助教授 / 准教授。2010年より現職。2018年より副所長。



■図-4 動作ユニット AI の社会実装例のイメージ