モーションキャプチャとステレオ映像を用いたロボットの遠隔操作

○松戸　康次郎*1　牧田　瞬*2
冨田　心太*1　澤　慎吾*3
入江　寿弘*4　新宮　清志*5

キーワード：ヒューマノイドロボット，モーションキャプチャ，両眼立体視，遠隔操作

1.　はじめに

近年，ロボットに関する研究や開発は，数多くの企業や研究所で行われており，特にヒューマノイドロボットにおいては，近年における半導体の技術革新により飛躍的な進歩を遂げている．また，ロボットの実用化に必要なそれぞれの技術においての発展は目覚ましく，現在においても，ディープラーニングによる会話の自動応答や機器の遠隔通信環境，ネットワークにおける情報の充実などによって，それらに関連する技術における研究や開発は盛んに行われ，今後もさらなる発展が期待される．しかし，それら高機能なロボットは，多くは研究などの用途に使用され，実社会で利用されているロボットでは簡単な会話や動作を行う程度にとどまっている．その大きな理由として，姿や形，それに付随する必要な機能を人型に収めたまま，低価格で人に近い精度や運動機能を備えることが困難だということである．ここでいう精度とは，行動や動作の合理性やそれに伴う安全性など，実用化するにあたっての必要な性能のことである．また，それらの問題を克服したロボットを実現するためには，多大な開発費や時間が必要であり，現時点では，日常的に利用することが困難であるといえる．

そこで，本研究では，現状の技術で利用可能なものを統合し，医療・介護現場や教育，エンターテイメントなどの場面で稼働することのできるヒューマノイドロボットの開発を目指す．

ロボットのモーションにおける制御には，再現性，同時性が高く容易にロボットを制御することが可能なモーションキャプチャシステムを用いた制御方法を採用した．

また，ロボットにモーションキャプチャシステムを導入するにあたり，視覚方面でもより再現性を高めた操作が行えるよう，VR ヘッドマウントディスプレイを用いたステレオ映像による両眼立体視システムを導入することにした．

これらの統合されたシステムは，ハードウェアとして，同研究室で開発しているエンターテイメント向け等身大キャラクターロボットである HRI-2[1)]に搭載した．

また，前述した遠隔操作システムのみでは活用できる場面が限られてしまうため，より汎用的に活用するためには，あらゆる機能の自動化が必要となる．そこで，マルコフ連鎖を利用し，特にエンターテイメントなど不特定多数の相手とコミュニケーションを図ることを想定した自動会話応答システムを構築することにした．

2.　研究目的

本研究は，姿勢制御・遠隔操作システム，VR 関連技術を用いたステレオ映像による立体視システム，不特定多数の相手との自動会話応答システムを構築し，実用化の観点から制御の難しいとされるヒューマノイドロボットに，これらのシステムを統合及び導入することを目的とする．

3.　モーションキャプチャシステム

モーションキャプチャとは，現実の人物や物体の動きをデジタル的に記録することのできる技術である．これをリアルタイムに通信することで，複雑な制御を必要としない直感的な全身の制御を可能にした．大掛かりな設備を必要としないため，計測場所に対する自由度が高く，対応する関節などの設定を済ませれば比較的容易に制御することができることから慣性センサ式のキャプチャ技術である Perception Neuron[2)]を使用することとした．

本システムは，操縦者の全身に取り付けられたモーションセンサから推定した角度が，統合開発環境内蔵型ソフトウェア Unity[3)]の 3D モデルと同期され，そのモデルにおけるそれぞれの関節角度情報を，実機であるロボットのサーボモータに入力することでロボットに操縦者と同様の動きをさせるという構成になっている．

4.　テレイグジスタンス

このシステムは，VR 関連技術を用いている．VR とは Virtual Reality の略称であり，日本語では，一般に「仮想現実(感)」と訳されている．世間で一般的に VR と言われているものは，実写やコンピューターグラフィックスで作り出した 360 度の映像の中にユーザーが入り込んだかのような没入感のある視聴覚体験ができるものを指している．その体験をするにあたり現在主流になっているのは，HMD(Head Mounted Display)を用いた手法である．本研究では，HTC 社の Vive[4)]を使用し実現した．

本来，体験者の動きを仮想空間に反映するためには，HMD の位置や向きを位置トラッキング技術などによって検出しなければならない．しかし，本研究においては前述

したモーションキャプチャによって操縦者の頭部の動きをロボットに反映させ，ロボット頭部に搭載したカメラでロボット周辺の視覚情報を HMD に映し出すことで，遠隔地にいる操縦者がロボット周辺の状況を把握すると同時に，ロボットを操縦することができる Telexistence(テレイグジスタンス：遠隔臨場感，遠隔存在感)システムを実現している．仮想空間における VR と同様に，実空間で自分自身がキャラクターとして演じることができる画期的なシステムと言える．

また，本研究で用いたロボットの構造上，搭載しているカメラが両眼に対応していないため，1 つの映像を両眼の視差分ずらし，左右それぞれのディスプレイに表示することで疑似的な立体視映像を投影していたが，後述のように両眼立体視についても検討を行った．

5. 音響システム

音声認識や音声合成のソフトウェアを利用し，双方向音声通信システムを構築した．

各デバイスの通信における一方向の指向性を考慮し，システム全体で聴覚情報を循環させる．構成として，操縦者の音声をキャラクターの音声として変換しロボットに発生させ，その音声も含めロボット周辺の音を操縦者に伝える．このようにシステム全体として通信を循環させることによってそれぞれの情報が混線せず，ハウリングすることもなく遠隔地での双方向音声通信が可能となる．

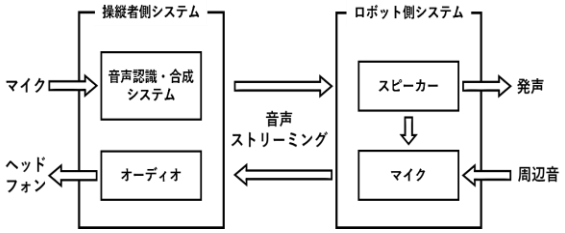

図-1　双方向音声通信システム

6. システムの統合

システムの全体像として，各システムとロボットのデバイスとの関係を図-2 に示す．図-2 において，モーションキャプチャと姿勢制御システムは姿勢・位置情報を通信し，その情報をもとにサーボモータへ入力する．また，ロボット頭部に搭載するカメラは，映像をストリーミングとして操縦者が装着している HMD に送信しており，音響システムについては前項に示した通りである。これらすべてのシステムを，Unity を経由して通信することで，モーションキャプチャシステム，音声認識及び音声合成を内包した音響システム，テレイグジスタンスシステムを同時に展開することができる．また，本研究では試験稼働のためデータを通信するのではなく，ロボットに搭載する小型デスクトップ PC(intel NUC7i7BNH)内の Unity に必要な機能及びシステムを集約させ，それをリモートデスクトップによって外部から操作することで実現している．

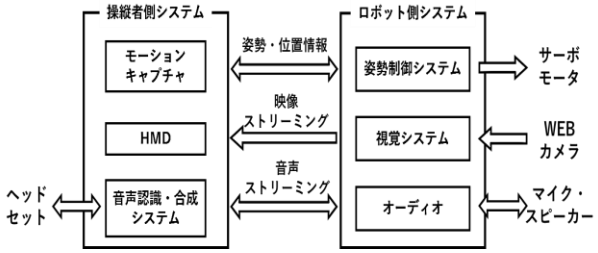

図-2　各システムとロボットのデバイスとの関係

7. ステレオ映像を用いた両眼立体視システム

(1)立体視

人間が立体感を感じ取る要因は多岐にわたる．両眼で機能するものとして，両眼視差，輻輳があげられ，単眼で機能するものとしては，焦点調節や運動視差などがあげられる．

仮想空間ではこれらの原理を利用し，仮想空間での視点から見た画像や映像に，人為的に作り出した視差や輻輳のずれを加えることによって立体的に感じさせている．

人間はある 3 次元物体を見たとき，特に両眼視差によって奥行きや距離感を感じ取っており，この両眼視差が立体感に与える影響が大きいため，その特性を利用し視差分だけずらしたステレオ映像を，それぞれに対応する観察者の眼に提示することで両眼立体視を可能にしている．また，その方法として HMD を用いることによって，視覚の立体化や位置トラッキングを容易に実現することが可能となった．

(2)両眼立体視システム

前述した通り，本研究におけるテレイグジスタンスシステムでは，使用したロボットの構造上の問題から，疑似的な立体視映像を投影していた．しかし，次世代機では両眼に対応させるため，ロボットの両目に搭載する 2 つのカメラによるステレオ映像を用いた立体視を可能とする両眼立体視システムを構築した．図-3 は実際に HMD に投影される映像の例である．

図-3　ディスプレイに表示される映像

8.　両眼立体視システムにおける最適値の計測

次世代機に搭載することを前提に，両眼立体視システムにおいてどのカメラ角度が，最も立体的認識の安定性があるか実験を行った．

実験装置として，正確なカメラ間隔と角度を設定するため，ベース・壁の役割を果たすケースに，カメラ間隔を変えることのできる平行移動用のサーボモータと，それぞれのカメラ角度を変えることのできる回転用のサーボモータを搭載し，立体的認識範囲の計測に用いた．サーボモータには，近藤科学の KRS-3204ICS[5]を使用した．

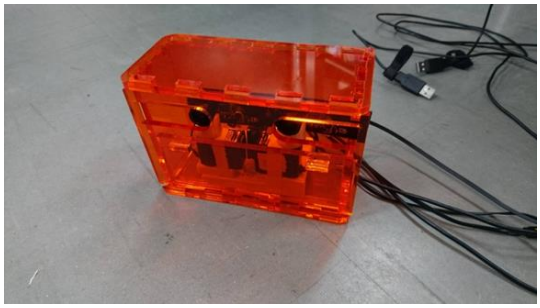

図-4　両眼立体視システム実験機

実験を行うにあたり，カメラ角度は互いに外向きとなる角度を正とし，最大値においてプロットのない領域は測定可能である目標物を立体的に認識することのできる範囲とする．

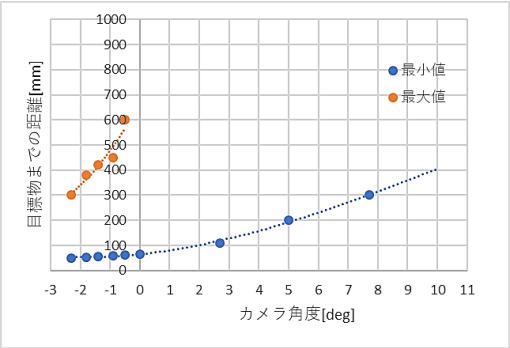

図-5　カメラ間隔 60mm

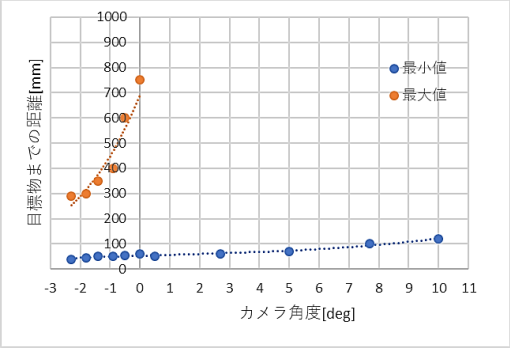

図-6　カメラ間隔 65mm

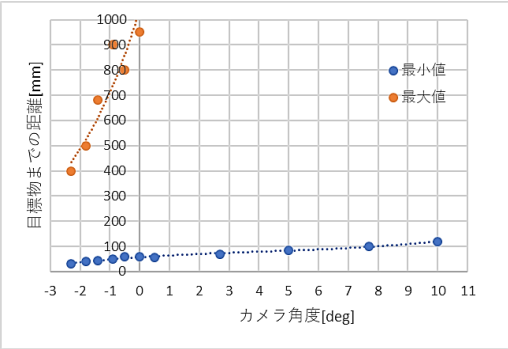

図-7　カメラ間隔 70mm

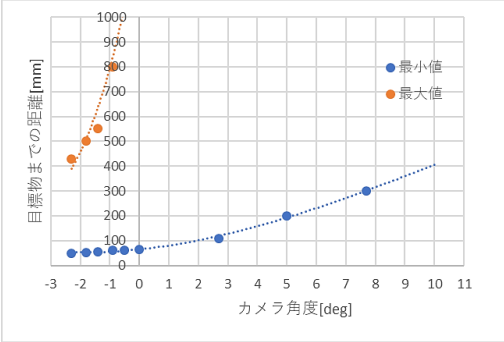

図-8　カメラ間隔 75mm

本実験の考察として，図-5 から図-8 を見ると各カメラ間隔において，立体的認識範囲の最小値はカメラ角度に対して大きな変動は見られなかった．人間の瞳孔幅に近しいカメラ間隔 65mm, 70mm では実際に見る人間の眼が補正しやすいため，ほかのカメラ間隔に比べ特にカメラ角度による最小値の変動が小さいことが確認できた．また，立体的認識範囲の最大値はカメラ角度 0[deg]近傍において遠方の値をとり，測定可能範囲の目標物が立体的に視認することができた．両眼立体視システムを実装するにあたり，ロボットの手元を含めカメラ位置から，少なくとも 150mm~遠方の立体的認識範囲が望まれる．

本実験において，前述の通り最も人間の瞳孔幅に近しいのは，カメラ間隔 65mm であるが，キャラクターとして瞳孔幅のバランスを考慮すると，カメラ間隔 75mm が望ましい．以上のことを踏まえ，次世代機に搭載するカメラの最適値は，カメラ間隔 75mm，カメラ角度+0.5[deg]であると推測される．

9.　n 次マルコフ連鎖を用いた自動会話応答システム

マルコフ連鎖とは，次の状態が過去の状態に依存せず，現在の状態のみによって決まる性質(マルコフ性)を持つ確率過程のうち，とりえる事象を示す値が離散的なものを指す．ここで，自然言語にマルコフ性を仮定すると，ある文章を形態素解析によって品詞分解された単語の羅列にし，その遷移確率から新しい文章を生成することで，入力元の文章らしい返答をすることができる．

例えば，夏目漱石の『吾輩は猫である』の冒頭部分数行を入力とし，マルコフ連鎖によって文章を生成すると，図-9 のような結果が得られた．

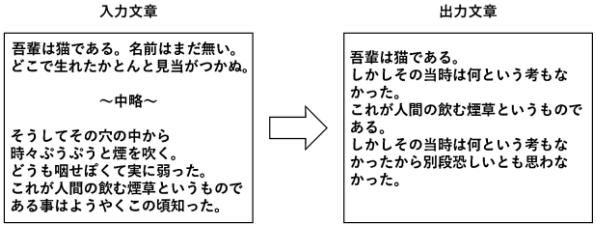

図-9　マルコフ連鎖による文章生成の例

図-9 を見ると，出力された文章は入力文章に似たような表現で，新たな文章が生成できることがわかる．この理論を適用した文章生成において性質上，人との対話で利用することは難しいと考えられる．しかしこの性質を，不特定多数を相手にしたエンターテイメントなどの場面に応用すれば，その状況に即したそれらしい返答を返すことや，また経験や法則に縛られすぎないため，人にはない発想の文章を生成することが期待できる．

また，本項のシステムを遠隔操作システムと同時に展開することで，より多様性のあるコミュニケーションを取ることができる．

10. 総括

人間が心理的にコミュニケーションをしていると感じることができているのは，仕草や表情，発声などを通し一連の言動に意味を見出していたり，自らの言動に対する応答や反応を認識していたりすることによると考えられる．従って，観客や応対者が違和感なくキャラクターを模したコミュニケーションロボットであると認識するために，必要なコミュニケーション方法として，四肢や表情の動き，キャラクター音声での発声が必要不可欠となる．本研究において，現時点でキャラクターを前提としたエンターテインメント向けのヒューマノイドロボットとして必要とする，最低限のシステムを同時に稼働させることができた．

よりキャラクターとして違和感なく認識させるためには，多くの部分を，自律し生物的に行動していると感じ取れることが重要となる．本研究のシステムでは操縦者が存在するため，それを認識してしまうと，あくまで命令に従うロボットであるという無機物的なイメージを感じ取ってしまうが，必要なシステムを内部に集約し，遠隔で操作することでそれを払拭することができた．

11. 今後の展望

現在開発中の次世代機では，本研究に基づきそれぞれのシステムの改善を検討する．モーションキャプチャシステムにおいては，自立や歩行に耐えうるサーボモータを脚部に搭載し，重心及びゼロモーメントポイントを算出することにより，自動的に姿勢を補正できる近藤科学の V-Sido OS[5]を併用することで歩行や移動の遠隔操縦を可能にし，本研究で用いた慣性センサ式のキャプチャシステムが苦手とする，指や表情のモーションを光学式などの他のキャプチャ技術との併用を検討する．また，遠隔で操作することによる遅延や処理速度による滑らかさの低下，地面との相対位置のずれなどの改善が必要である．

テレイグジスタンスシステムにおいて，立体的視覚情報は遠隔操作において重要な情報であるため，本研究で推定した最適値をもとに両眼立体視システムの導入が望まれる．両眼立体視システムの導入にあたり，それぞれのカメラ映像の FPS が同じであっても，表示タイミングのずれによる VR 酔いが考えられ，また，厳密には実際の眼球で見た場合とは異なる状態の映像を補正して，立体視をするために生じる視覚的疲労感，これらを軽減することも必要となる．

音響システムにおいては，返答の遅延がコミュニケーションにおける違和感を助長してしまう．そのため，リモートデスクトップを経由することによる遅延や，音声認識，音声合成の処理速度の改善が望まれる．また，遠隔操作を行う際，操縦者にとって立体的視覚情報と同様に重要であるのが，ロボット周辺の聴覚情報である．ロボットの周辺音をステレオ方式やバイノーラル方式を用い立体音響にすることで，視覚情報だけでは知覚できない情報を，正確に把握し補完することができる．

また，本研究で開発したようなロボットの遠隔操作は，エンターテイメント向けのロボットや介護・医療，教育現場などでの限定的な利用であれば有用性があるが，前述の通り今後，より日常的・汎用的に広く活用するためには，多くの部分の自律化が必要となる．それに伴い，より自然で高効率な 2 足歩行，ディープラーニングによるモーション生成や会話応答，画像認識，さらには，自己位置推定や障害物の検知などの技術を搭載することが考えられる．

[参考文献]

1) 冨田 心太，鈴木 圭太，山口 尚菜，入江 寿弘，新宮 清志：2 足人型キャラクターエンターテイメント・ロボットの開発 -リアリティと記号の中間の感情表現- 計算工学講演会論文集，Vol.22 2017.5
2) Perception Neuron：https://neuronmocap.com/ja
3) Unity：https://unity3d.com/jp
4) VIVE：https://www.vive.com/jp
5) 近藤科学：http://kondo-robot.com/

＊1　日本大学大学院　理工学研究科　精密機械工学専攻　院生
＊2　日本大学理工学部　精密機械工学科　学生(研究当時)
＊3　京都大学大学院　情報学研究科　知能情報学専攻　院生
＊4　日本大学理工学部　精密機械工学科　教授　博士(工学)
＊5　日本大学　名誉教授　工学博士，総合資格学院　特別顧問