

遠隔演奏におけるインタラクションを支援するための ジェスチャ交換システム

ファン ラック フック¹ 池上 智彦² 高田 秀志¹

概要: 通信技術の発展に伴い、遅延の極めて少ない通信環境での遠隔作業が可能になった。離れた場所にいる複数の音楽演奏者が、ネットワークを介してお互いの演奏を聴きながらリアルタイムに合奏する遠隔演奏もその一つである。このような演奏でも、対面演奏時の身振り手振りなどによる非言語インタラクションも同様にやり取りできるように支援する必要がある。本稿では、インタラクションの一種である身振り手振りによる意思疎通を支援するためのジェスチャ交換システムを提案する。本システムでは、各種のセンサにより、演奏者のジェスチャを検出する。ジェスチャが発生した場合、演奏相手のコンピュータにジェスチャ情報を送信する。このジェスチャ情報を受信したコンピュータでは、送信した演奏者の姿を3Dアバターとして再現し、表示する。受信側の演奏者は3Dアバターの動きから、送信側の演奏者の意思を理解し、それを考慮して適切に演奏を合わせる。検証実験の結果から、システムが遠隔演奏に対して有益であり、正確性・有用性・利便性があることが示唆された。

A Gesture-Exchange System for Supporting Interaction in Remote Ensemble

PHUC PHAN LAC¹ TOMOHIKO IKEGAMI² HIDEYUKI TAKADA¹

Abstract: The advancement of communication technology has paved way to perform remote works that require low latency communication connection. Remote ensemble is one of the examples that enable multiple musicians in remote locations to play ensemble in real time while listening to the performance of each other via network. In face-to-face ensemble, the musicians usually use body language to interact with each other for collaboration. This paper presents a gesture-exchange system to support interaction based on body language in remote ensemble. This system uses sensors to detect gestures of musician. When a gesture is detected, the gesture information will be sent to other musicians' computer. The computer that has received the gesture information will render the gestures by using 3D avatar of gestured musician. By watching the 3D avatars, the musician can understand each other's intention and take appropriate action. The experiment results have shown that this system has high accuracy, usefulness and convenience to play remote ensemble with multiple musicians.

1. はじめに

通信技術の発展に伴い、遅延の極めて少ない通信環境での遠隔作業が可能になった。離れた場所にいる複数の音楽演奏者が、ネットワークを介してお互いの演奏を聴きなが

らリアルタイムに合奏する遠隔演奏もその一つである。

一般的に対面環境で合奏する場合、お互いの演奏を合わせるために、楽器の音のほかにも身振り手振りによるインタラクションが演奏者間では行われる。ところが、従来のネットワーク技術では、マイクにより収録した楽器の音声のデータを送受信するとリアルタイム性を保ちながら演奏することが困難な程の遅延が発生したため、このような演奏環境は実現されていなかった。そのため、遅延問題を十分に解決する方法のなかった今までは、音声データの再現

¹ 立命館大学情報理工学部
College of Information Science and Engineering, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan

² 立命館大学大学院情報理工学研究科
Graduate School of Information Science and Engineering, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan

のみ重要視され、演奏者間のインタラクションはあまり考慮されていなかった。しかし、近年のネットワーク通信の高速化に伴い、遅延時間が 30ms まで低減されたため、遠隔演奏を行えるようになった。本研究では、このような遠隔演奏を行う場合に、演奏者同士のジェスチャのやり取りを相互に行うことで、対面環境での合奏に近いインタラクションを実現し、遠隔演奏を支援する。

このような環境を実現するには、演奏の支援を行う上で、より多くの遅延を発生させないことが必要とされている。本研究は、音声データ以外のネットワーク負荷を可能な限り抑えた合図のやり取り方法を利用したシステムとして、ジェスチャ交換システムを提案する。各演奏者は、複数の加速度センサ・ジャイロセンサを身体に装着し、1台の情報端末を手もとに用意する。システムは加速度センサによって演奏時のジェスチャを検出し、ジェスチャが発生した場合、演奏相手の情報端末にジェスチャ情報を送信する。ジェスチャ情報を受信した演奏相手の情報端末には、受信したジェスチャ情報に基づき、その演奏者の姿を 3D アバターとして再現し、表示する。演奏者は、3D アバターの動きから何の合図が送信されたかを確認し、演奏相手の意思を理解できたとき、その意思に従って演奏を合わせる。本システムでは映像の送受信と比べ、ネットワークの負荷が低くジェスチャの検出率が高いという利点がある。さらに、3D アバターの動きが対面演奏時の実際のジェスチャと類似するので、演奏者の臨場感を向上させることができる。

本研究では、上記のシステムを構築し、評価実験を実施する。実験の結果から、ユーザがどのようにシステムを活用できるかを調査し、システムが遠隔演奏に対して有益であるかを検証する。

2. 遠隔演奏

2.1 遠隔演奏環境

遠隔演奏とは、離れた場所にいる複数の音楽演奏者が、ネットワークを介してお互いの演奏を聴きながらリアルタイムに合奏することである。

このような演奏環境を構築するとき、前提条件として遅延の問題を解決しなければならない。許容できる遅延時間を計測するために、対面環境において実験が行われた [1]。この実験では最初に、演奏者 2 人が背中合わせとなり、お互い離れながら合奏をした。その結果、10m (往復 20m) まで離れた場合、合奏が混乱し、続けることができなくなってしまった。その 2 人の距離と音速 340m/s から計算すると、30ms (往復 60ms) となった。これにより、30ms 以上の遅延時間が発生する環境では、合奏することができないという結論が得られた。

ネットワークで音声を送受信するとき、ネットワーク通信や音声のエンコード・デコードによる遅延が発生する。対面演奏における遅延問題と同様に、上記の各要素により

発生した遅延時間の総計が 30ms を超えると、遠隔で合奏を行うことができない。

従来のネットワーク通信において、遅延の問題を解決するため、MIDI データを用いた手法が提案されることが多かった。MIDI データは電子楽器の出力である軽量データなので、演奏者間で MIDI データをやり取りすることで、遅延時間をあまり発生させないことができる。しかし、歌手の声を再現することができないことが MIDI データの欠点である。また、演奏者にとって、生楽器の音を収録した高音質なオーディオを聞くほうが楽しいという事実がある。当時のネットワークにおいて高音質なオーディオをやり取りすると、遅延時間が多発するため、混乱なく演奏を合わせることができていなかった。

近年、通信技術の進歩により、ヤマハの「NETDUETTO」[2] を活用した遠隔地間を繋ぐ技術と、NTT 東西の「フレッツ 光ネクスト」および「ひかり電話」の帯域確保および優先制御機能 [3] とを活用し、低遅延な演奏環境が実現されている。この環境では、インターネット越しに、離れた場所にいる演奏者同士での生楽器を用いた音楽セッションを行うことができるようになった。このような技術を用いたサービスをユーザに提供したところ、遅延時間 30ms 以下でのセッションが可能であった。このことから、遅延の問題は十分に解決できたと言える。今後は、ネットワークを介する遠隔音楽演奏がさらに受け入れられると考えられる。

しかし、前提条件として上記の遅延問題を十分に解決できていない頃には、ほとんどの研究は演奏者同士の意思疎通や臨場感を考慮せず、ただ遅延問題のみに注目していた。遠隔で演奏できるにとどまらず、演奏を豊かにできるように対面演奏環境と類似する遠隔演奏環境を提供することはこれから注目すべき課題となった。その課題を解決するためには、演奏者間のインタラクションを支援することが必要となる。

2.2 遠隔演奏におけるインタラクション

遅延問題は 2.1 節に述べたとおり解決されたが、遠隔においても豊かに演奏できる環境を構築するために、もう一つの重要な要素としてインタラクションを支援する必要がある。対面演奏では、演奏者間の様々なインタラクションが存在するが、身振り手振りに基づいた合図交換は代表的なものの一つである。特にジャズセッションには、このような非言語インタラクションが活用される。インタラクションは演奏時の意思疎通のためであり、演奏の豊かさをもたらしすることができる。

例えば、図 1 に示すような身振り手振りによる非言語インタラクションを活用した対面演奏がよく見かけられる。最初に演奏者 A は「ソロしたい」という意思を演奏者 B に伝えるために、言葉で表現する代わりにうなずいている。演奏者 B はそのジェスチャを確認し、演奏者 A の意思を

理解できた場合、演奏者 A をソロさせるようにする。

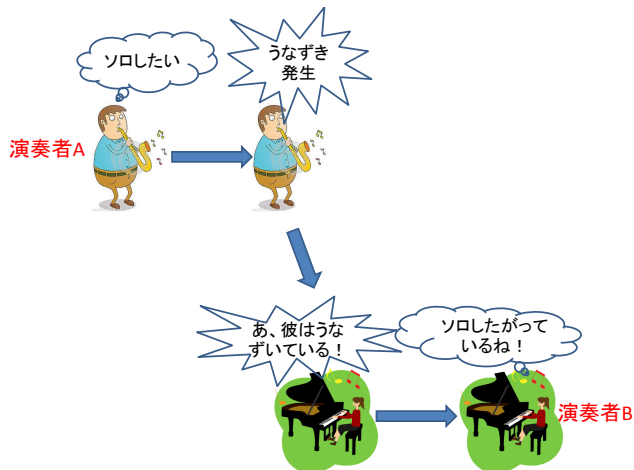


図 1 対面演奏におけるインタラクションの一例

しかし、遠隔で合奏する場合、お互いの姿と身振り手振りが見えないため、演奏者は自分の演奏意思を伝えることができない。そこで、意思疎通の実現できる環境を提供できるように、既存の遠隔演奏システムを拡張する必要がある。それに伴い、演奏の豊かさと演奏者の臨場感を向上させることもできると思われる。

3. ジェスチャ交換システム

3.1 システムの全体像

本節では、演奏者間の身振り手振りによる非言語インタラクションの必要性を考慮し、それを支援するシステムを提案する。このようなシステムを構築するとき、一番簡単なのは演奏者の姿の映像を収録してネットワーク経由で生中継する手法である。しかし、映像のような大容量なデータを送受信することで、ネットワークの負荷になる。その場合、解決できた音声遅延問題が再び発生し、音声のやり取りにも影響を与え、合奏が困難になってしまう。そこで、撮影した映像データを生中継する手法の代わりに、音声の遅延問題を再発生させない軽量なデータで演奏者の身振り手振りを伝達する手法を利用する。意思疎通を目的とした身振り手振りをジェスチャとして送信し、受信側で再現して表示する手法はその一つである。

この手法でシステムを構築するときの構成を図 2 に示す。ジェスチャ検出部として演奏者の身体に複数のセンサを装着したり、ジェスチャ再現表示部として演奏者の傍にコンピュータの画面を配置したりする。

本システムでは、ジェスチャ交換は図 3 に示すように行われる。仮に、2 人の演奏者 A, B がいるとする。演奏者 A は自分の意思を伝えたいと思うとき、ジェスチャを発生させる。そのジェスチャがジェスチャ検出部により検出され、ジェスチャの種類情報が得られる。その種類情報がネットワークを通して演奏者 B のコンピュータへ送信さ

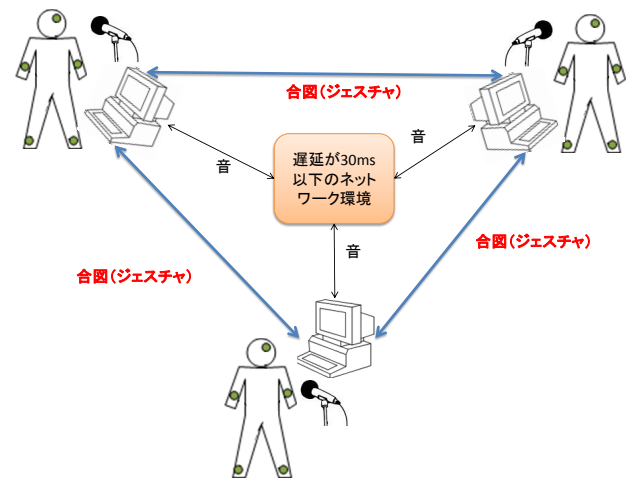


図 2 システム構成

れる。演奏者 B のコンピュータでは、受信したデータからジェスチャの種類情報を取得し、それに基づいて画面に演奏者 A のジェスチャを再現し、表示する。演奏者 B は表示しているジェスチャに気づいてから、演奏者 A の意図が理解でき次第、それに対応して演奏を合わせる。

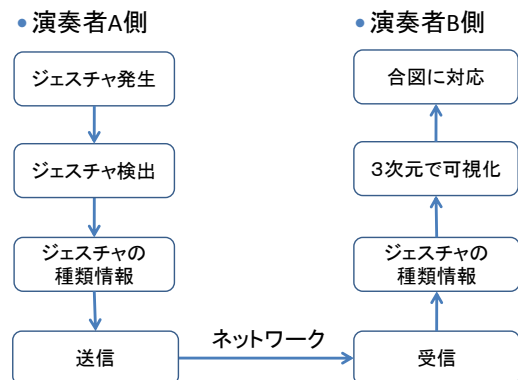


図 3 ジェスチャ情報交換の手順

映像データに比べ、ジェスチャの種類情報は軽量なデータであり、ネットワークを通じて送信しても遅延問題が発生しない。

このようなシステムを構築するために、ジェスチャ検出機能とジェスチャ再現表示機能を構築する。

3.2 ジェスチャ検出

ジェスチャ検出部は複数の加速度センサとジャイロセンサから構成される。表 1 に示すように、検出すべきジェスチャの特徴によって、演奏者の身体に装着する位置と適用するセンサの種類が選ばれる。主に加速度センサが活用されるが、首を左右へ振るジェスチャのような回転運動を検出するにはジャイロセンサが必要である。

本システムでは、加速度センサとジャイロセンサが付いている iPod Touch などの iOS デバイスを用い、ジェスチャ検出部を構築した。

表 1 ジェスチャに応じるセンサの種類

ジェスチャ	位置	センサの種類
うなづく	頭	加速度センサ・ジャイロセンサ
首を左へ振る		
首を右へ振る		
左手をあげる	左手	加速度センサ
左手を振る		
右手をあげる	右手	加速度センサ
右手を振る		
左足を踏む	左足	加速度センサ
右足を踏む	右足	加速度センサ

3.3 ジェスチャ再現表示

各演奏者のコンピュータには、ジェスチャ種類情報を受信し、そのジェスチャを 3D アバターとして再現して表示するためのプログラムがある。このプログラムは、3D ゲーム開発のためのゲームエンジンである Unity3D を用いて実現されている。

図 4 に示すように、各演奏者は画面で自分の視線から演奏相手の姿を観察できる。構築したシステムではギター、ピアノ、バイオリンを弾く 3 人の合奏を支援できる。

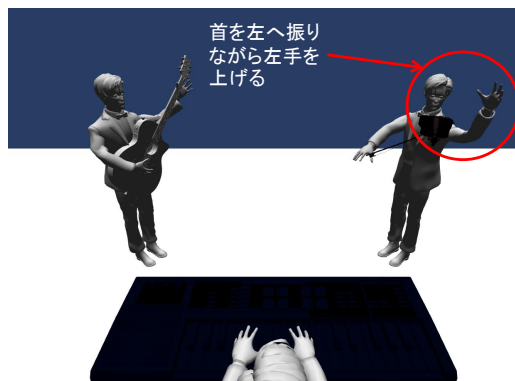


図 4 ピアノ演奏者の視線から見る再現画面

ジェスチャが発生したことを確認できたとき、そのジェスチャに込められる演奏相手の意思を理解できた上で、それにふさわしい演奏で対応することができる。また、演奏者は演奏相手のジェスチャだけでなく、自分のジェスチャもこの画面で見ることができ、正確に検出できたかを確認できる。

4. 実験および評価

4.1 実験

システムの正確性、有用性を検証し、また、システムが演奏者によってどのように実践的に利用されるかを調査することによって、遠隔演奏におけるインタラクションを支援できるかを確認することを目的として実験を行う。

4.1.1 実験の内容

検証実験は、2 年にわたって対面演奏で身振り手振りをお互いにやり取りする経験がある 3 人の被験者で実施す

る。各演奏者にはそれぞれ別の部屋においてギター、ピアノ、バイオリンで演奏してもらう。1 つの部屋の環境を図 5 に示すように設定する。遅延時間が 30ms 以下の環境を提供するために、この実験ではネットワークを介さずに、各演奏者のマイクとスピーカをアナログ機材でつなぎ、音声を取り取りさせる。

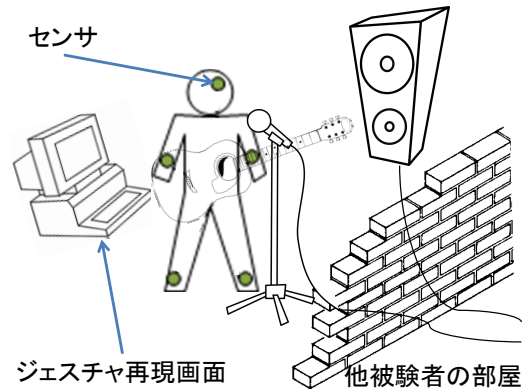


図 5 実験環境

実験環境を前述のとおり設定し、各演奏者にクリスマスメドレー (Silent night, First Noel, Joy to the world のメドレー) を演奏してもらう。バイオリン演奏者はメロディを担当し、ピアノ演奏者とギター演奏者が伴奏する。

演奏している時、システムにより以下の情報が記録される。

- ジェスチャが発生した時点
- ジェスチャの種類
- 発生した演奏者

また、システムにより記録された情報に基づいて各演奏者に以下のような質問をする。

- ジェスチャが発生した演奏者への質問
 - (1) 時点 t で発生したジェスチャは誰に向けましたか?
 - (2) その時点 t ではどのような演奏状況でしたか?
 - (3) そのジェスチャの意味は何でしたか?
- ジェスチャを受けた演奏者への質問
 - (1) 時点 t で発生した相手のジェスチャに気がつきましたか?
 - (2) 演奏相手の意思に対応する演奏を合わせてあげましたか?

以上の手順に沿って実験を実施した後、記録情報および各演奏者の回答に基づき、実験の結果をまとめる。

4.1.2 結果

実験によって、図 6 に示すようなジェスチャ交換の結果が得られた。矢印はどの演奏者からどの演奏者へジェスチャが投げられたかを示す。それぞれの矢印に付けられている数字はジェスチャの番号である。ジェスチャの番号と表 2 を照らし合わせることで、ジェスチャの種類とそれに込められた意思が分かる。ジェスチャ情報を受けた演奏者

がそのジェスチャに対して反応せずに無視した場合、そのジェスチャでは意思疎通できなかったと判断し、有用ではないと見なす。

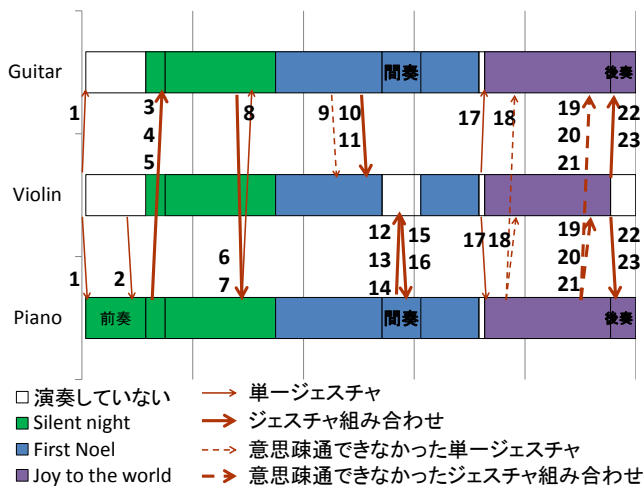


図 6 合奏の流れおよびジェスチャ交換

表 2 発生したジェスチャの意味

番号	ジェスチャ	ジェスチャの意味	有用
1	うなづく	Silent night の前奏を開始しよう	○
2	右手をあげる	Silent night の前奏を終了させよう	○
3	首を左へ振る	盛り上げすぎないように	○
4	左手をあげる		
5	左手を振る		
6	首を右へ振る	First Noel へ遷移しよう	○
7	うなづく		
8	うなづく	First Noel への遷移を了解	○
9	右手をあげる	間奏を入れて一人で演奏したい	X
10	右手をあげる	間奏を入れて一人で演奏したい	○
11	右手を振る		
12	右手をあげる		
13	左手をあげる	First Noel のエンディング準備しよう	○
14	首を右へ振る		
15	首を左へ振る		
16	うなづく	エンディング準備の合図を了解	○
17	うなづく	全員あわせて Joy to the world の演奏を開始しよう	○
18	左手をあげる	なし	X
19	右手をあげる	なし	X
20	左手をあげる		
21	左手をあげる		
22	左手をあげる	メロディが終わったので伴奏も終えよう	○
23	右手をあげる		

一人の演奏者から連続して発生したジェスチャは一つの意味のみを伝える目的で他の演奏者に送られる傾向があったため、ジェスチャ組み合わせとしてまとめた。ジェスチャが発生した 23 回の中で、意味が「なし」と表記された 18, 19, 20, 21 番は演奏者が意図的に行ったジェスチャではなく、システムの誤検出によって記録されたものである。

4.1.3 考察

演奏者がどのようにジェスチャ交換システムを利用してインタラクションを行ったかを理解するために調査を行う。実験結果により以下のことが分かる。

- 足によるジェスチャ交換機能が利用されなかった。その理由は一般的な対面演奏でも演奏者同士はお互いの足を観察し、足で発生したジェスチャに気づくことが難しいので、演奏者同士の意思疎通のために足の運動が利用されていないからである。
- 同じジェスチャの種類であれば、異なる演奏状況の下で発生してもジェスチャで伝える意思が類似する。表 2 を基にして分析すると、表 3 が得られる。

表 3 ジェスチャの意味に関する共通点

ジェスチャの種類	伝達意思の共通点
うなづく	開始の合図あるいは受けた合図に賛同
首を左右へ振る	ジェスチャを受ける対象者を指定
手をあげる	まもなく終わりの合図
手をあげて振る	やめてほしいと伝えたい

伝達意思の共通点が多く見られた理由は、各演奏者が人間の一般的な身体言語を活用したことである。例えば、うなづくことで合意している意思を表現したり、首を左右へ振ることで向きを示したり、手を振ることで何か否定したりするような身体言語を日常生活でよく見かける。そのような身体言語を活用する場合、言葉で表現しなくても、人間同士はお互いに理解し合えることができる。音楽演奏で発生した身体言語もその例外ではないと考えられる。

- ジェスチャ組み合わせを投げるのはあいまいさを避けるためである。上記のように身体言語が利用されるが、身体言語そのものがあいまい性を持つ。例えば、右手を上げるジェスチャである 9 番のジェスチャをギター演奏者がバイオリン演奏者へ投げた。しかし、そのジェスチャだけではあいまいすぎて、バイオリン演奏者が「間奏を入れ、一人で演奏したい」という意思を理解できず、演奏しているメロディを止めなかった。意思疎通が失敗した後、やり直すために、ギター演奏者が 10-11 番のジェスチャ組み合わせを投げた。9 番のジェスチャと比べると、これは更なる複雑なジェスチャ組み合わせであり、演奏者の意思をより具体的に反映させるので、意思疎通が実現できるようになった。
- 以上により、ジェスチャ交換システムは意思疎通に基づいた演奏者間のインタラクションを支援でき、遠隔演奏に役立つという結論が得られる。

4.2 評価

システムの効果を以下の 3 つの要素で評価する。

- 正確性

実験の結果としてジェスチャが 23 回発生したとシステムに記録されているが、4 回が誤検出であるので、19/23 (82.6%) が正確に検出されたことになる。

その 4 回の誤検出は Joy to the world を演奏している最中に起こった。この曲はテンポが速くて、盛り上がる曲であり、ピアノ演奏者が非常に激しく演奏したことが理由で、手に付いた加速度センサの提供する加速度データが大きく変化し、検出を誤った。

これにより、遅く演奏するときにはシステムが正確に検出できるが、速くて盛り上がる曲を演奏するとき、その正確性を保証できないと評価できる。

- 有用性

ジェスチャ交換が演奏者間の意思疎通に役立つかを基準として、システムの有用性を評価する。ここでは、合図を受けた演奏者が演奏相手の意思に答えてあげる回数を有用性の指標にする。システムの誤検出が理由で発生したジェスチャは、システムで送られても、それを受けた演奏者が理解できないし、意思疎通に悪影響を与えるので、有用ではないジェスチャとして扱う。実験結果のとおり、9 番のジェスチャ、18 番のジェスチャと 19-20-21 番のジェスチャ組み合わせが有用ではなかった。ただし、ジェスチャ組み合わせは一つの意思のみを表現するので、複数回のジェスチャ交換として扱わずに、一回のジェスチャ交換と同様に扱う必要がある。そうすると、ジェスチャ交換が意思疎通に役立たない回数は 3 回である。即ち、ジェスチャ交換の有用率は 10 回/13 回 (76.9%) である。

- 利便性

本システムは利便性の面で大きく利点がある。

ユーザがどのように意思疎通できたかを考察した結果によると、ユーザにとって束縛がなく、日常生活の身体言語でインタラクションを行う環境ができたことである。したがって、システムの使い方が分かりやすく、短期間でユーザ習熟度を向上させることができるのは一つの利点である。その上に、ユーザが自由にシステムを活用して豊かな遠隔演奏ができると、対面演奏と類似する臨場感をもたらしことができる。その他に、ユーザは自分の発生したジェスチャを画面で確認できるおかげで、ジェスチャ検出機能が正確に動作しているかをすぐ確認できることがもう一つの利点である。しかし、構築したシステムにはまだ制限点が残存する。例えば、複数のセンサを身体に装着したり、ネットワーク通信のために手動で各端末の IP アドレスを指定したりする必要があるため、利用事前設定が複雑であり、ユーザにとって面倒になるという欠点がある。

以上の評価より、速い曲を演奏していてもジェスチャ検出機能が正確に動作するように改良したり、環境設定の複雑さを下げたりすることが今後の課題として挙げられる。

そのような課題を解決するために、利用するセンサの種類とセンサ数を再検討する必要がある。また、手動で IP アドレスを設定するかわりに、仮想の合奏ルームで相手にマッチングする機能を導入していくことで、利便性を向上させるように検討する。

5. おわりに

本研究では、遠隔演奏における非言語インタラクションを支援することを目的とし、ジェスチャ交換システムを構築した。映像データをやり取りするのに比べ、本システムでは、軽量のデータをやり取りすることで、すでに解決された遅延問題が再び発生しないという利点がある。構築したシステムの効果を検証するために実験を実施し、演奏者がどのようにシステムを利用してインタラクションを行ったかを調査した。その結果、ユーザが身体言語を通して、演奏者同士の意思疎通ができることが判明した。演奏者はお互いの意思を考慮して合奏の流れを合わせることで、より豊かな演奏ができるだけでなく、対面演奏時と類似する臨場感も感じられるようになった。このことから、本システムが遠隔演奏に対して有益であることが分かる。また、本システムには正確性・有用性・利便性があることが示唆された。しかし、利用する前の環境の設定がまだ複雑であり、今後改良のための検討を行う必要がある。

参考文献

- [1] 後藤 真孝, 根山 亮: "Open RemoteGIG: 遅延を考慮した不特定多数による遠隔セッションシステム", 情報処理学会論文誌, Vol.43, No.2, pp.299-309, 2002.
- [2] ヤマハ株式会社: インターネットを介してリアルタイムで、遠隔地間の合奏・重唱などを可能にする技術「NET-DUETTO(ネットデュエット)」を開発 (online), 入手先 <<http://www.yamaha.co.jp/news/2010/10100501.html>> (参照 2014-1-30) .
- [3] 西日本電信電話株式会社: フレッツ光を利用した新しい音楽スタイルを実現する「ひかり D U E T T O NY 1」の販売開始について (online), 入手先 <<http://www.ntt-west.co.jp/news/1211/121121b.html>> (参照 2014-1-30) .
- [4] 後藤 真孝, 日高 伊佐夫, 松本 英明, ほか: "仮想ジャズセッションシステム: VirJa Session", 情報処理学会論文誌, Vol.40, No.4, pp.1910-1921, 1999.
- [5] 赤澤 慶一, 田邊 浩之, 星加 百合絵, ほか: "音楽ライブにおける遠隔地ファンとエンタテイナーのアニメーションを用いたコミュニケーション支援", エンタテインメントコンピューティングシンポジウム 2013 論文集, pp.234-237, 2013.
- [6] Masahiro Y., Yuka O. and Tatsuhiro Y.: *A Protocol For Remote Musical Session with fluctuated tempo*, Proceedings of the 2004 International Conference on Cyberworlds, pp.87-93, 2004.
- [7] Shengzhong Y., Yongzhao L. and Hong H.: *MIDI-based Software for Real-time Network Performances*, 2010 First ACIS International Symposium on Cryptography, and Network Security, Data Mining and Knowledge Discovery, E-Commerce and Its Applications, and Embedded Systems, pp.226-230, 2010.