

ニッポン放送賞

VRにおける観客アバタの提示手法についての
包括的なデザインガイドラインの開発

筑波大学大学院 システム情報工学研究群
博士後期課程2年

矢倉 大夢

1. 序 論

音楽ライブの鑑賞はバーチャルリアリティ (VR) の有望な応用の1つである。例えば、Facebookは2018年にOculus Venuesというプラットフォームを立ち上げ、ビリー・アイリッシュ等の著名なアーティストのライブにVRで参加することを可能としている。また、VTuberと呼ばれる、3Dアバタをモーションキャプチャで操作する新たな形態のキャラクター (Shirai, 2019) たちがVR環境でライブを行うというケースも広まりつつある。一方で、そうしたリモートでのライブ参加には乗り越えるべき課題が未だ存在するという指摘もある。例えばTarumiら (2017) は、リモートでのライブ参加は観客間でリアクションを共有することで得られる「一体感」に欠くとしている。これはWebbら (2016) の、コプレゼンス (Co-Presence) —他者と共にある感覚—が、観客をパフォーマンスに惹きつけ、ライブ感 (Liveness) を生み出すために重要であるという指摘とも一致する。

こうした議論を元にTarumiら (2017) は、よりよいライブ参加体験を実現するためにアーティストのパフォーマンスだけでなく観客間でリアクションも共有されるようなシステムを開発する必要があると結論付けている。実際Oculus Venuesのような既存プラットフォームも、ヘッドマウントディスプレイ (HMD) から取得されるユーザの動きをVRライブ会場内の観客のアバタに反映するという機能を実装している。しかしこの機能は、それなりの数のユーザがHMDを用いてVRプラットフォームに接続し、同時に同じライブを鑑賞するという点を暗に仮定している。言い換えると、もしVRライブを鑑賞しているユーザがかなり少なかったり、あるいは、1人しかいなかったりした場合に、この機能のみでは上述のようなライブ参加のソーシャルな体験を味わうことはできない。

逆に言うと、もし観客アバタのモーションをコンピュータによって自動的に提示することができれば、そうした状況にあるユーザにもよりよいライブ参加体験を提供できる可能性がある。さらに、観客アバタのモーションを自動提示するというアプローチは、VR環境での没入体験を実現するという点に関してライブ参加以外の用途にも応用できる。例えば、VR空間において映画館のような環境で動画を鑑賞できるビデオプレーヤーアプリケーションが広まりつつあるが、これに観客アバタの自動提示を組み合わせることで、観客がペンライトを振ったり、踊ったりしながら大人数で映画を鑑賞するという「応援上映 (英語圏ではsing-alongsとも)」 (Allen, 2011; Alexander, 2014; Tomoko, 2016) の体験をVR上で多くのユーザに提供できると考えられる。

そこで、本論文では観客アバタの自動提示を行うVRライブ鑑賞システムを提案する¹。ただし、観客アバタをこうした形で活用するという点についてこれまでの研究がないため、こういった仕組みでその動作を提示するのがよいか、様々な手法を包括的に比較検討するところから行った。具体的には4つの提示手法を提案し、それらを組み込んだVRライブ鑑賞システムを実装してそれぞれの有効性を検証するとともに、その結果を元にVRライブのよりよい参加体験を生み出すためのガイドラインを開発した。

2. 背 景

本章ではまず、音楽ライブ鑑賞へのVRの応用の状況について概説するとともに、それらの課題について議論する。また、観客アバタのモーションを提示するという点に関連して、

¹ 本論文の内容の一部はIEEE ISMAR 2020にて行った発表 (Yakura and Goto, 2020) に基づいている。

アバタモーションの合成に関する既存研究を紹介する。

2.1 VRによるライブ鑑賞

VRのエンタテインメントへの応用は長年にわたって行われてきた(Sherman and Craig, 2002)。その中でも、VR技術によって生み出されるプレゼンス(Presence) —自らが「そこにいる」という没入した感覚(Slater and Wilbur, 1997; Usoh et al., 2000) —は、ユーザの生の感情を引き出す上で特に重要と考えられてきた(Riva et al., 2007)。特に、コンピュータを通した画面越しのパフォーマンスはプレゼンスを実現できていないという指摘もあり(Diemer et al., 2015)、多くの研究者がVR技術を活用したライブ参加システムを開発してきた(Geigel, 2017; Kaneko et al., 2018; Atarashi et al., 2018; Kasuya et al., 2019; Horie et al., 2018)。

VRのエンタテインメントへの応用という点でもう1つ挙げられるのがCinematic VR (MacQuarrie and Steed, 2017)である。これは、VR空間内において映画館のようなパノラマビューで映像を鑑賞するという体験を指しており、冒頭で述べたようにCinematic VRのためのビデオプレーヤアプリケーションも数多く存在している。2025年にはVRによるライブ参加の市場規模は41億ドルに達するとされており(Bellini et al., 2016)、VRによるライブ鑑賞は産業的にも有望視されている応用の1つであると言える。

2.2 VRによるライブ鑑賞の課題

そうした期待の一方で、乗り越えるべき課題もいくつか指摘されている。Tarumiら(2017)は半構造化インタビューの結果から、ライブへのリモート参加は現実世界での参加体験と比較して一体感を欠いていると指摘している。ここでの「一体感」とは、他の観客と一体となってパフォーマンスへとリアクションする中で得られるもので、観客間の相乗的な感動を作り出すのに重要であるという。このライブ参加のソーシャルな側面の重要性を強く物語っている例として、「初音ミク」のライブ(Jones et al., 2015; Leavitt et al., 2016; Lam, 2016)を挙げることができる。初音ミクはバーチャルキャラクターであるがそのライブが世界中で開催されており、会場ではキャラクターのパフォーマンス映像が投影されたスクリーンの前で、多くの参加者が音楽に合わせペンライトを一体となって振っている(Lam, 2016)。このパフォーマンス自体は映像コンテンツに過ぎないにも関わらず観客が会場に詰めかけている理由として、Leavittら(2016)は他の観客と一緒に参加することで生まれる相乗的な感動と、そこから得られるライブ感にあると分析する。

こうした点からいくつかの研究(Tarumi et al., 2017; Webb et al., 2016)が、遠隔でのライブ参加における一体感やコプレゼンスの醸成のために、観客間でリアクションを共有できるシステムの必要性を議論している。実際、そうした機能は既存のVRライブ鑑賞プラットフォームにも実装されており、それぞれのユーザの身体の動きがVR空間内の観客アバタにリアルタイムに反映されるようになっている。しかし冒頭でも議論した通り、このアプローチはそれなりの数のユーザが同時に接続して鑑賞していることを仮定しており、使える状況は限られている。さらに、人は音楽を聞いているときにビートからずれて動いている相手を見るとネガティブに感じる(Stupacher et al., 2017)ことを考えると、ネットワークの問題等によって他のユーザの動きの反映が遅れてしまった場合に、ユーザ体験を損なう恐れもある。こうした点からVR環境での観客アバタの提示手法について包括的な比較検討が求められる。

2.3 アバタモーションの自動合成

ライブパフォーマンスに合わせて観客アバタのモーションを提示するという問題は、楽曲に合わせて時系列的な人の動作を生成するという点で、ダンス振り付けの生成 (Ofli et al., 2012; Fukayama and Goto, 2015; Tang et al., 2018) に関連するという点もできる。ただし、これらは基本的にダンサー1人の表現的な振り付けを生成することを目的としており、ライブにおける観客のように多くのアバタがシンクロしたモーションを見せる様子を提示したいという目的に直接用いることはできない。複数のアバタのモーションを生成するという点では、群衆シミュレーション (Thalmann and Musse, 2007; Xu et al., 2014) も近いモチベーションを持っている。しかし、これらの手法は音楽に合わせて生成するという点は考えておらず、音楽とのシンクロが重要となるライブの観客アバタのモーションの提示には用いることができない。以上から、VRライブで観客アバタのモーションを提示するという目的のためにダンス振り付けの生成や群衆シミュレーションの既存手法を適用するのは難しいと考えられる。そうした点でも、本論文で述べる包括的な比較検討の重要性が示唆される。

3. 実装とリソース

VRライブにおける観客アバタのモーションの提示手法の比較検討のために、我々はまず Oculus Quest (HMD) を用いてVRライブ鑑賞システムを実装した。このシステム上では、ユーザが観客アバタと共にVR空間内の仮想ライブ会場にいるかのように提示される。観客アバタの数は200体以上で、それらの動きは提案手法のいずれかによって自動制御される。

また、VRライブの環境を再現するにはライブパフォーマンスの提示も必要なため、初音ミクの3Dモデルをパフォーマンスさせる仕組みも実装した。これは、2.2章で述べたように、初音ミクのライブは他の観客の存在がライブ参加の大きなモチベーションになっているという点で、提案手法の応用シナリオとして妥当であるという点が挙げられる。さらに、初音ミクをテーマにした創作コミュニティがインターネット上に形成されており (Leavitt et al., 2016; Lam, 2016)、楽曲がアップロードされているだけでなく、それら楽曲に合わせて初音ミクの3Dモデルがダンスするモーションを公開しているクリエイタも存在する (Leavitt et al., 2016; Hamasaki and Goto, 2013)。そのため、そうしたパブリックなリソースをVRライブでのライブパフォーマンスの再現のために用いることができるという理由もある。

また、そうしたリソースを利用しながらファンが自分たちでプロジェクタとスクリーンを用意し、初音ミクのライブを開催するということが行われており (Leavitt et al., 2016)、そのライブの様子もインターネット上に公開されている。そのライブ映像を利用することで、2.3章で紹介したダンス振り付けの生成手法と同じように、観客アバタのモーションをデータドリブンに合成することが可能となる。特に、初音ミクのライブにおいては観客の存在が大きな意味を持つ (2.2章) ことから、これらのライブ映像でも観客の様子がしっかりと映されているという特徴があり、一般的なアーティストのライブ映像と比較しても有用である。

まとめると、図1のような形で初音ミクの3Dモデルによるライブパフォーマンスを周囲

の観客アバタと共に鑑賞できるVRシステムを実装した。また、提示するライブパフォーマンスとして、ニコニコ動画で公開されている楽曲とそのダンスモーションを5曲分収集した。合わせて、ファンによるライブ3回(2回は日本で、もう1回はアメリカで開催)から38曲分の映像(合計2時間54分)を収集した。

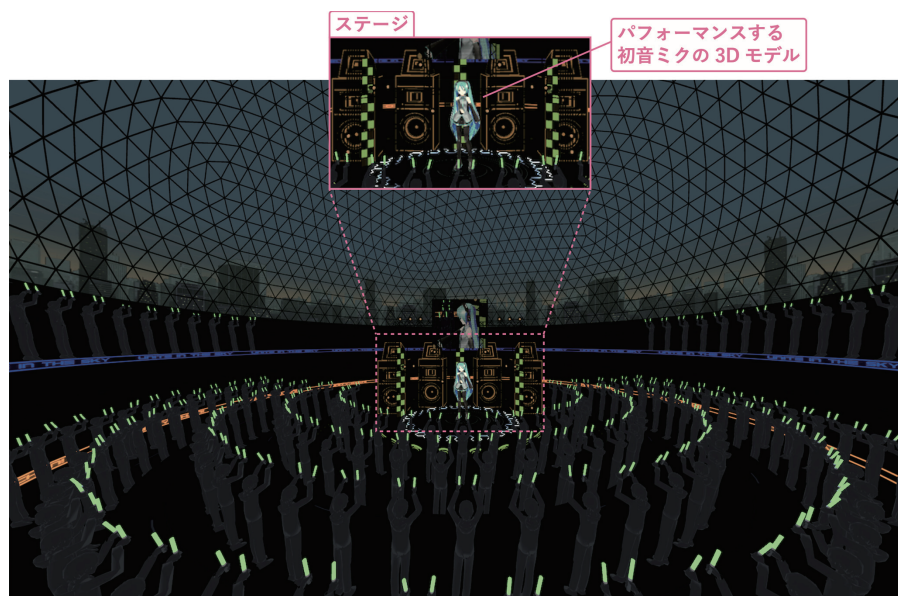


図1：実装したVRライブ鑑賞システムによる表示の例。ユーザーは、中央のステージでパフォーマンスする初音ミクの3Dモデルと、現実世界のコンサートにおける観客と同様にペンライトを持っている周囲の観客アバタの両方を見ることができる。

4. 提案手法

本節では、観客アバタのモーションを楽曲に合わせて提示するために3章で述べたVRライブ鑑賞システムに実装した4つの手法について詳説する。まず、ユーザ自身あるいは他のユーザの動きをコピーして提示するという2つのアプローチを、引き続いて、楽曲の情報を活用して動きを提示する2つのアプローチを導入する。

4.1 ユーザ自身の動きのコピー

1つ目のアプローチとしては、ユーザ自身の動きをそのまま観客アバタにコピーするという手法を考える。これにより、ユーザが音楽にシンクロして動いている限り、観客アバタのモーションも音楽に合ったものとなる。この手法は観客アバタのモーションの提示のために追加のリソースを必要せず、実装も非常に容易である。

4.2 他のユーザの動きのコピー

次に、同じコンテンツをすでに鑑賞した他のユーザの動きを観客アバタにコピーするという手法を考える。具体的には、システム上でそれぞれのユーザの動きをライブパフォーマンス上の時間軸(開始後何秒経過したか)とともに記録する。そして、他のユーザが同じコンテンツを鑑賞したときに、観客アバタがそのタイミングで過去に他のユーザが取っていた動き

を再現する。この手法の利点は、現実世界のライブに他のユーザと同時に参加したのとほぼ同じような状況を再現できるため、自分の動きをそのままコピーする場合よりリアリティが増すと期待できる点にある。

これは、オンライン動画共有サービスにおける「弹幕コメントシステム」(Goto, 2011; Wu et al., 2019) に着想を得たものである。これにより、それぞれのユーザが現実の時間上では全く異なるタイミングに動画を視聴していたとしても、動画を見ている限りは他の多くのユーザと同時に視聴しているかのような感覚を得られる。この「疑似同期」と呼ばれる体験は、ユーザのコプレゼンスを高める (Ma and Cao, 2017; Chen et al., 2017) だけでなく、ユーザにライブ感を生み出す (Johnson, 2013) と知られている。

4.3 ビートに同期したモーションの繰り返し

続いて、ユーザの動きをコピーするのではなく、アバタのモーションをコンピュータで合成するというアプローチも実装した。具体的にはまず、楽曲中のビートに同期してモーション素材を繰り返すという手法を実装した。これは、3章で収集したライブ映像を分析したところ、ほとんどの観客が限られたパターンの動きをビートに合わせてながら小節ごとに繰り返しているという様子が観察できたためである。

ここでは、データドリブンにモーション素材を用意すべく、まず観客の代表的な動きのパターンを現実世界のライブ映像からクラスタリングアルゴリズムで抽出することとした。クラスタリングアルゴリズムへの入力としては、ビート検知情報 (Goto, 2006) を元に小節ごとに分割したライブ映像から、観客のペンライトの動きを抽出して用いた。動きの数値化には、人物行動認識で広く用いられる特徴量である HOF (Histogram of Optical Flow) (Laptev et al., 2008) を用い、その結果に時系列 k-平均法 (Petitjean et al., 2011) を適用することで4つのクラスタを得た。図2は得られた4つのクラスタそれぞれについて、クラスタの中心に最

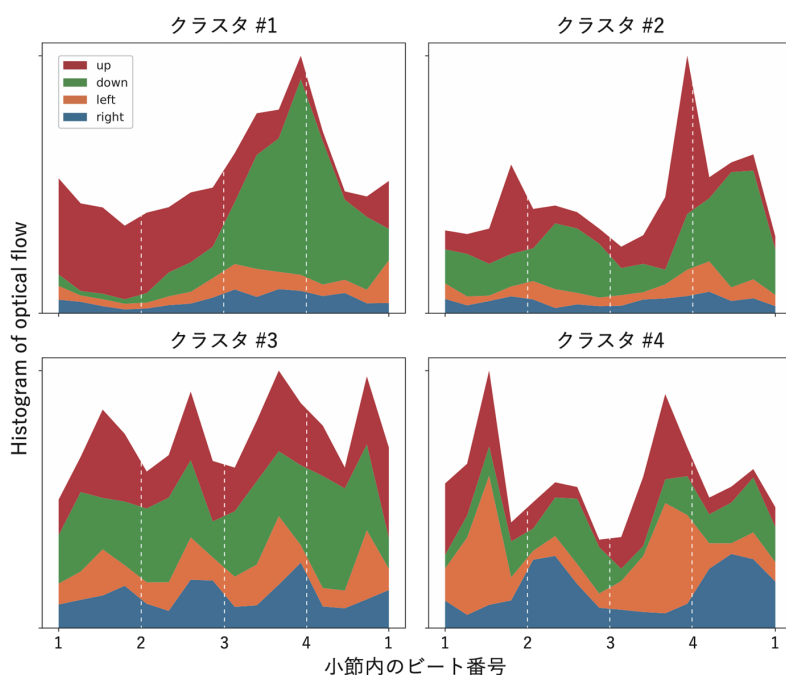


図2：4つのクラスタのそれぞれの代表的な小節内のHOF。白い点線はビートのタイミングを表し、対応する動きは図3の通りである。これら HOF は手の相対的な動きを示しているため、拍の直前にピークが現れる。

も近かった代表的な小節での観客のペンライトの動きを可視化したものである。そして、ここに示した小節内でのHOFの上下左右への動きを元に観客の動きを図示したものが図3である。この4つのクラスタに対応する結果が、ライブでの観客の動きを経験的に分類した既存研究でも述べられている (Mori et al., 2017) ことを考えると、これら4つの動きをモーション素材として用いるのは妥当であると考えられる。

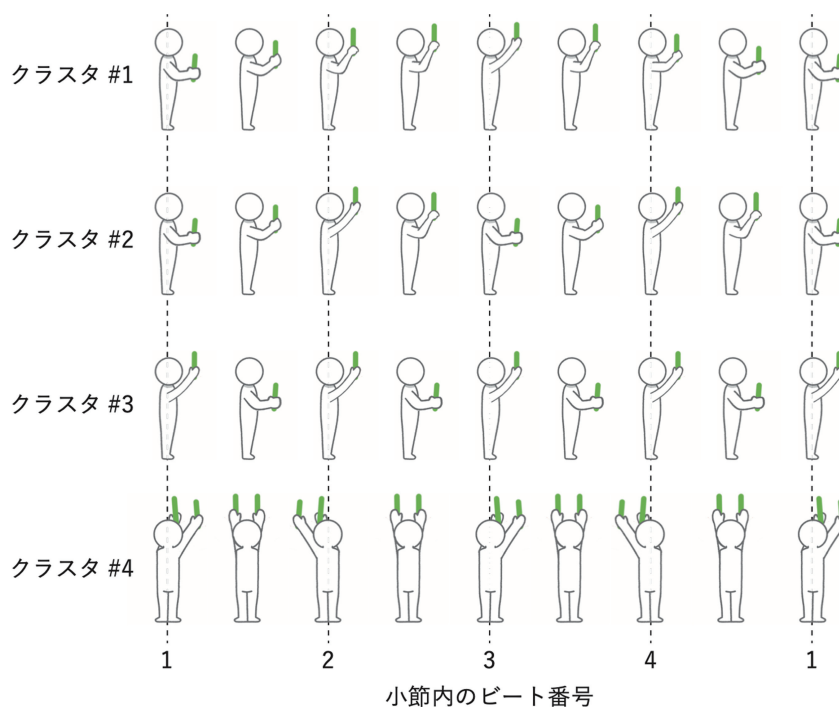


図3：図2の4つのクラスタに対応する代表的なモーションのイラスト。白い点線は小節内のビートのタイミングを表している。

以上の結果に基づいて、前述のようにモーション素材を小節ごとに観客アバタに繰り返させる手法を実装した。具体的な実装としては、楽曲が始まる度に観客アバタそれぞれに、これら4つの素材のうちの1つをランダムに割り当てるようにした。ただしその分布は、表1に示した現実世界でのライブ映像における4つのクラスタの出現割合に対応するようにした。そして、ビート検知アルゴリズムで検出された小節のタイミングに基づいて、観客アバタが割り当てられたモーション素材を繰り返すようにした。

表1：現実世界でのライブ映像における4つのクラスタの出現割合。拍ごとに手を上げる動き (クラスタ #3) が最も支配的であった。

クラスタ	#1	#2	#3	#4
出現回数	276	682	1,132	207
出現割合	12.0%	29.7%	49.3%	9.0%

4.4 機械学習による合成

4つ目の手法としては、機械学習を用いて曲調に合ったモーション素材を選び出し、楽曲ごとに適したモーションを合成するというアプローチを採った。これは、3章で収集したライブ映像をさらに分析したところ、観客が楽曲の雰囲気に合わせて代表的なパターンの動きの中から無意識的に適したものを選び出し、動きを切り替えている様子が観察されたためである。つまり、この曲調と動きの関係を機械学習で適切にモデル化できれば観客の動きをよりリアルに再現できるということになるため、まず事前実験としてそうしたモデル化が可能かを検証することを行った。

機械学習モデルに入力する特徴量としては4種類を用意し、それらの組み合わせを変えながらLightGBM(Ke et al., 2017)で精度を比較した。そのうち3種類はTzanetakisら(2002)が機械学習による楽曲のジャンル分類のために提案したもので、音色特徴量、リズム特徴量、ピッチ特徴量からなる。残り1種類は、楽曲中の繰り返し区間やサビ区間の位置を示した大域構造特徴量(Goto et al., 2011)であり、現実世界のライブ映像においてサビ区間ではより活発な動きを選ぶなど、観客が大域的な楽曲構造に影響を受けている様子が観察できたため加えた。結果は表2に示す通りで、いずれの特徴量の組み合わせについても4.3章の分布からランダムサンプリングした場合の期待値より高い精度が得られた。最も高い精度を得られたのが4つすべての特徴量を組み合わせた場合で、いずれの特徴量も精度向上に寄与していることが確認された。

表2：4つの特徴量の組み合わせごとの観客の動きの予測精度。すべての特徴量を用いた場合が最も高い精度を示した。

特徴量				精度
音色	リズム	ピッチ	大域構造	
✓				69.35%
	✓			62.87%
		✓		62.17%
			✓	64.21%
✓	✓			70.70%
✓		✓		70.70%
✓			✓	74.18%
	✓	✓		66.57%
	✓		✓	70.27%
		✓	✓	69.31%
✓	✓	✓		72.18%
✓	✓		✓	74.27%
✓		✓	✓	73.97%
	✓	✓	✓	71.01%
✓	✓	✓	✓	76.67%
ランダムサンプリングによる期待値				35.36%

以上の結果に基づき、機械学習を用いながら楽曲に合わせて観客アバタのモーションを提示する手法を実装した。この手法では、それぞれの小節に対して図3の4つの動きの尤度を計算し、その尤度に対応する割合となるようにそれぞれの観客アバタにモーションを割り当てる。そして、その小節の間、観客アバタは割り当てられたモーション素材(4.3章と同一)を再現する。

5. 比較実験

4章で述べた4つの手法について、その有効性を比較し、VRライブにおいて観客アバタの提示を提示することの効果を検証するためにユーザ実験を行った。本節ではその手順と結果を詳説する。

5.1 尺 度

本論文の冒頭や2.2章で述べたように、観客アバタの提示手法を検討する背景には、現在のVRライブが一体感やコプレゼンスをユーザに提供できていないという課題がある。言い換えると、観客アバタのモーションを提示することでユーザの一体感やコプレゼンスを高められたかを検証する必要がある。ただし、Tarumiら(2017)の言う一体感を計測するための確立された尺度は存在していない。またVR環境内でのコプレゼンスについても、これまでテレプレゼンスシステムの評価などで用いられてきた尺度(Slater et al., 2000; Biocca et al., 2001; Nowak and Biocca, 2003)はあるが、人と人の間のインタラクションの質を計測するものとなっているのに対し、提案手法ではユーザと観客アバタの間での明示的なインタラクションはもたらされないため、適さない。

そこで、この比較実験ではHwangら(2015)のコプレゼンスの尺度を翻案して用いた。これは、スポーツのライブ中継をテレビで視聴する際にオンラインコミュニケーションを組み合わせることで生まれるコプレゼンスを計測するという目的で開発されたものである。弾幕コメントシステムがユーザに与える影響を検証した研究(Cameron, 2016; Chen, 2018)でも用いられていることから、コンテンツ鑑賞のためのインタラクションの評価という点で提案手法の効果の検証に用いるのも妥当であると言える。また、VRライブ鑑賞の総合的な没入感を評価するために、VRシステムに関する既存研究(Riva et al., 2007; He et al., 2018)と同様にユーザのプレゼンスも計測した。ここでは、VR環境におけるプレゼンスの尺度として特に広く用いられている尺度の1つである(Schwind et al., 2019)、SUSスケール(Usch et al., 2000)の6つの質問を用いた。

5.2 手 順

比較実験は、それぞれの参加者が異なる楽曲と提示手法の組み合わせからなる2回のライブ鑑賞体験を行う設計とした。参加者は20名(20歳から46歳、うち5名は女性)で、ペンライトを用いてライブに参加した経験のある参加者を募ったため、男女比についてはそうしたライブ文化の同質性(To, 2014)を反映したものとなっている可能性がある。提示手法としては、4章で述べた4つに加え、観客アバタをモーションなしで提示するベースラインを含めた5つの中から2つを選び、参加者の楽曲への嗜好度や順序効果を抑制するような形で割り当てた。



図4：実験の実施の様子。参加者はOculus Questを着用し、開発したVRライブ鑑賞システムを体験した。

5.3 結果

図5はそれぞれの提示手法に対する参加者のコプレゼンスの評価を示したものである。Durbinの検定 (Durbin, 1951; Conover, 1999) によって有意な差があることが確かめられた ($D(4) = 12.84$, $p = 0.01$) ため事後検定を行ったところ、モーションなし及び自らの動きのコピーの場合と、それ以外の3つの提示手法の場合との間で有意な差が確認された ($p < 0.05$)。

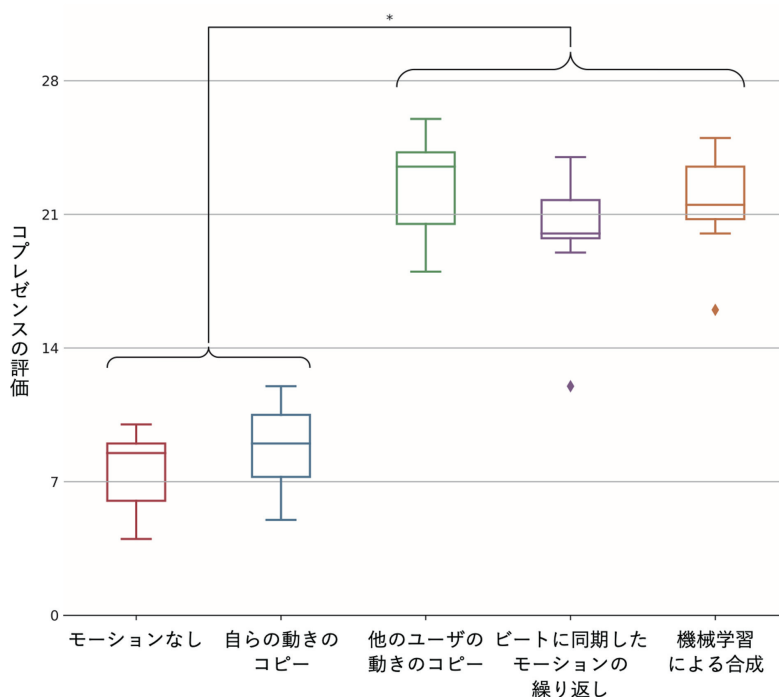


図5：それぞれの提示手法に対する参加者のコプレゼンスの評価。モーションなし及び自らの動きのコピーの場合が、他の3つの場合と有意に異なる結果 ($p < 0.05$) を示した。

図6はそれぞれの提示手法に対する参加者のプレゼンスの評価を示したものである。同じく Durbin の検定 ($D(4) = 10.22, p = 0.04$) を経て事後検定を行ったところ、他のユーザーの動きのコピー及び機械学習による合成は、モーションなし及び自らの動きのコピーの場合との間に有意な差を示す ($p < 0.05$) ことが確かめられた。また、参加者ごとにコプレゼンスとプレゼンスの評価に高い相関がある (スピアマンの $\rho = 0.52, p = 0.001$) ことも分かった。

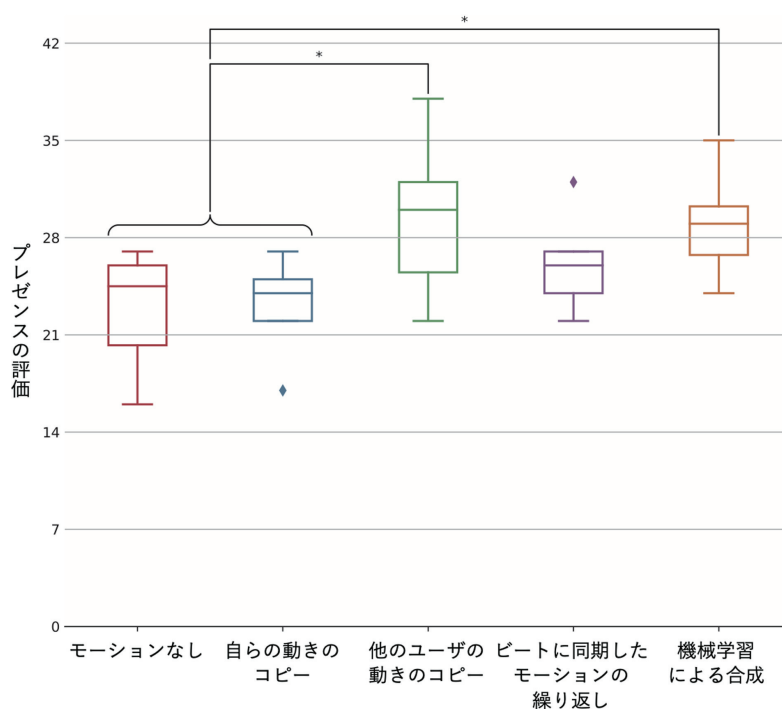


図6：それぞれの提示手法に対する参加者のプレゼンスの評価。モーションなし及び自らの動きのコピーの場合が、他のユーザーの動きのコピー及び機械学習による合成の場合と有意に異なる結果 ($p < 0.05$) を示した。

さらに、参加者それぞれに経験した2つの提示手法のどちらがよい体験だったかを聞き、Plackett-Luceモデルで (Maystre and Grossglauser, 2015) 相対的な評価の数値化を行った。結果は表1の通りで、これは5つのうちから2つを取り出して比較したときに、どちらが高い評価を受けるかという確率がこのスコアの比となることを示している。この結果は、特に他のユーザーの動きのコピー及び機械学習による合成が高い評価を受けたという点で、コプレゼンスやプレゼンスの評価と似た傾向を示している。また、自らの動きのコピーが、モーションなしの場合より低いスコアとなっているのも興味深い結果である。

表3：Plackett-Luceモデルで算出されたそれぞれの提示手法のスコア。

提示手法	スコア
モーションなし	0.085
自らの動きのコピー	0.020
他のユーザーの動きのコピー	0.387
ビートに同期したモーションの繰り返し	0.140
機械学習による合成	0.368

5.4 分 析

5.3章で示した通り、自らの動きをコピーした場合を除いて、提案手法はユーザのコプレゼンスを高める効果があった。自らの動きのコピーに対する評価が低かった理由は、当該の手法を体験した参加者の「自分の動きがコピーされていることにすぐ気づき、アバタがクローン人間みたいに見えてきて少し気持ち悪かった」というコメントから推察できる。逆に言うと、他のユーザの動きをコピーした場合には、観客アバタのモーションに実際の観客たちのようなばらつきが見られることから、高い評価につながったのではないかと考えられる。

同時に、他の3つの提案手法は参加者によりよい鑑賞体験を提供するのに貢献し、特に他のユーザの動きのコピー及び機械学習による合成は、ユーザのコプレゼンスとプレゼンスを有意に高める効果があると確認された。他のユーザの動きのコピーが高評価だったのは、4.2章で述べたように他のユーザと一緒にパフォーマンスを鑑賞しているのとはほぼ同じような状況が再現できるためであると考えられる。一方、ビートに同期したモーションの繰り返しによる提示と機械学習による合成による提示は同じモーション素材を用いているにもかかわらず、その評価に違いがあった理由としては、機械学習を用いることで曲調に合わせて自然なモーションの切り替えができたことによると言える。

加えて、コプレゼンスとプレゼンス、そしてPlackett-Luceのモデルによるスコアの間で似たような傾向が見られた。これは、一体感やコプレゼンスがよいライブ参加体験を生み出すのに重要であるという点を質的手法で指摘した既存研究 (Tarumi et al., 2017; Webb et al., 2016) の結果を、量的に補強するものとなっている。

6. 議 論

5章での比較実験を通して、観客アバタの提示がVRでのライブ参加体験の向上につながることが示唆された。本節ではその結果を元に、こういった状況でどの提示手法を用いるべきか、デザイン上の指針を議論する。また、提示手法の実装や比較実験の結果に関する制限についても明らかにする。

表4：VRライブにおける観客アバタの提示に関するガイドライン。これらは5章の実験結果を基に整理されたものである。

	ライブストリーミング	非公開コンテンツの再生	パブリックなコンテンツの再生
自らの動きのコピー	非推奨	非推奨	非推奨
他のユーザの動きのコピー	(低遅延の通信回線が保証されていない限り) 非推奨	適用不可能	(モーション情報が入手可能であれば) 強く推奨
ビートに同期したモーションの繰り返し	推奨	適用可能	適用可能
機械学習による合成	(大域構造特徴量なしでモデルを学習すれば) 適用可能	推奨	推奨

6.1 デザイン上の指針

比較実験の結果から、自らの動きをコピーする以外の提示手法はVRライブへの参加体験の向上に有用であると言える一方で、残りの3つの手法はモーション提示のためにそれぞれ異なるリソース要件を持つ。そうした点から、比較実験の結果をベースに応用シナリオを議

論し、観客アバタを提示する上でのデザイン上の指針(表4)を示す。なお、提案手法を用いるシナリオとしてはインターネット上でのライブストリーミング、パブリックなコンテンツの再生、非公開コンテンツの再生の3つとした。現在のVRライブ鑑賞プラットフォームは主にライブストリーミングにのみ対応しているが、2.1章で述べたようにCinematic VRの形でライブ映像やその他の動画コンテンツを再生するという応用も含めるためである。

ライブストリーミングにおいては、他のユーザの動きをコピーして提示するとネットワークのレイテンシによって動きが音楽からずれてしまい、逆に体験を損なってしまう可能性があるとして2.2章で述べた。その点で、5低遅延のインターネットアクセスが保証されていない限り、他のユーザの動きをコピーする手法は推奨されない。また、ライブストリーミングにおいては次にどんな楽曲が再生されるのかを前もって知っておくことができないため、機械学習による合成を行うにしても大域構造特徴量を計算できないという課題がある。そのため、機械学習による合成の方が高い評価を受けたとはいえ、楽曲の音響信号からリアルタイムで抽出できる(Müller et al., 2011)ビート情報のみを用いて、モーション素材を繰り返すという提示手法が妥当であると言える。

ただし、表4では機械学習の入力として大域構造特徴量を用いなかったとしても、モーションの選択の精度が大きく落ちることはなかった。残り3つの特徴量は小節単位で計算できるものとなっているため、ライブストリーミングをバッファリングしてこれら特徴量を計算するように実装を変更することで、機械学習による合成を適用することは可能となる。逆に、手元にある非公開コンテンツを再生するような状況を考えると、大域構造特徴量を含めすべての特徴量を事前に計算しておくことができる。さらに、非公開コンテンツに対する他のユーザの動きのモーション情報を入手するのは困難であることを考えると、機械学習による合成が妥当な選択と言える。

パブリックなコンテンツの再生に関しても、機械学習による合成を適用することは可能である。一方、オンライン動画共有サービスでユーザのコメントが共有されている(4.2章)ように、パブリックなコンテンツに対するユーザの動きを共有するようなサービスが近い将来に生まれるという可能性は否定できない。そうした時代が訪れれば、パブリックなコンテンツの再生の際に他のユーザの動きをコピーして観客アバタを提示するということも可能である。

6.2 制 限

比較実験から観客アバタの提示の有効性が確認され、デザイン上の指針を得ることもできた一方で、ここまでの議論にはいくつかの制限が存在する。例えば、結果の一般性を担保するためにはより多様でジェンダーバランスの取れた参加者を募り、大規模な追加実験を行うことが望ましい。また、5.1章ではコプレゼンスとプレゼンスを尺度として提示手法の比較を行うことにしたが、ライブ参加体験の様々な側面を明らかにするために他の尺度を組み合わせることも可能である。さらに、参加者の楽曲への嗜好度や参加者自身の動きの大きさが最終的な評価と関連を見せるのかを分析することで、さらなる発見を得られる可能性もある。

7. まとめ

本論文では、ユーザのコプレゼンスを向上させ、よりよい参加体験を生み出すために観客

アバタのモーションを自動提示するVRライブ鑑賞システムを導入した。具体的には、自らの動きのコピー、他のユーザの動きのコピー、ビートに同期したモーションの繰り返し、機械学習による合成の4つの提示手法を提案し、それらの有効性を比較実験することを行った。実験結果からは、他のユーザの動きをコピーして提示する手法が参加体験の向上に非常に有効であると分かった一方で、機械学習による合成も同程度に有効かつ多くのシナリオに適用可能な選択肢となることが明らかとなった。これらの結果は、ユーザの意識の周縁にアバタのモーションを自動提示することでユーザ体験を向上させるというインタラクションデザインの有用性を支持するものである。

謝 辞

本論文で述べた研究の一部は、科学技術振興機構 ACT-X(JPMJAX200R)、ACCEL(JPMJAC1602)及び日本学術振興会科学研究費(JP21J20353)の支援を受けている。また、著者はGoogle Ph.D. Fellowship及びMicrosoft Research Ph.D. Fellowshipに選出され、研究支援を受けた。

参考文献

- Akihiko Shirai. REALITY: broadcast your virtual beings from everywhere. In *Proceedings of the 46th ACM SIGGRAPH Conference – Appy Hour*, pages 5:1–5:2, New York, NY, 2019. ACM.
- Hiroyuki Tarumi, Tomoki Nakai, Kei Miyazaki, Daiki Yamashita, and Yuya Takasaki. What do remote music performances lack? In *Proceedings of the 9th International Conference on Collaboration Technologies and Social Computing*, pages 14–21, Cham, Switzerland, 2017. Springer.
- Andrew M. Webb, Chen Wang, Andruid Kerne, and Pablo César. Distributed liveness: Understanding how new technologies transform performance experiences. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 431–436, New York, NY, 2016. ACM.
- Robert C. Allen. Reimagining the history of the experience of cinema in a post-movie-going age. *Media International Australia*, 139 (1): 80–87, 2011.
- Bryan Alexander. Sing-along ‘Frozen’ coming to theaters. USA TODAY, <https://www.usatoday.com/story/life/movies/2014/01/22/frozen-sing-along/4783327/>, January 2014.
- Kawabata Tomoko. New movie showings give audiences chance to cheer, sing along. The Mainichi, <https://mainichi.jp/english/articles/20161023/p2a/00m/0et/006000c>, October 2016.
- Hiromu Yakura and Masataka Goto. Enhancing participation experience in VR live concerts by improving motions of virtual audience avatars. In *Proceedings of the 19th IEEE International Symposium on Mixed and Augmented Reality*, pages 555–565, Washington, DC, 2020. IEEE.
- William R. Sherman and Alan B. Craig. *Understanding Virtual Reality: Interface, Application, and Design*. Morgan Kaufmann, Burlington, MA, 2002.
- Mel Slater and Sylvia Wilbur. A framework for immersive virtual environments five: Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6 (6): 603–616, 1997.

- Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, 9 (5): 497–503, 2000.
- Giuseppe Riva, Fabrizia Mantovani, Claret Samantha Capideville, Alessandra Preziosa, Francesca Morganti, Daniela Villani, Andrea Gaggioli, Cristina Botella, and Mariano Alcañiz. Affective interactions using virtual reality: The link between presence and emotions. *CyberPsychology & Behavior*, 10 (1): 45–56, 2007.
- Julia Diemer, Georg W. Alpers, Henrik M. Peperkorn, Youssef Shiban, and Andreas Mühlberger. The impact of perception and presence on emotional reactions: A review of research in virtual reality. *Frontiers in Psychology*, 6: 26, 2015.
- Joe Geigel. Creating a theatrical experience on a virtual stage. In *Proceedings of the 14th International Conference on Advances in Computer Entertainment Technology*, pages 713–725, Cham, Switzerland, 2017. Springer.
- Tatsuyoshi Kaneko, Hiroyuki Tarumi, Keiya Kataoka, Yuki Kubochi, Daiki Yamashita, Tomoki Nakai, and Ryota Yamaguchi. Supporting the sense of unity between remote audiences in VR-based remote live music support system KSA2. In *Proceedings of the 1st IEEE International Conference on Artificial Intelligence and Virtual Reality*, pages 124–127, New York, NY, 2018. IEEE.
- Ray Atarashi, Takuro Sone, Yu Komohara, Manabu Tsukada, Takashi Kasuya, Hiraku Okumura, Masahiro Ikeda, and Hiroshi Esaki. The software defined media ontology for music events. In *Proceedings of the 1st International Workshop on Semantic Applications for Audio and Music*, pages 15–23, New York, NY, 2018. ACM.
- Takashi Kasuya, Manabu Tsukada, Yu Komohara, Shigeki Takasaka, Takuhiro Mizuno, Yoshitaka Nomura, Yuta Ueda, and Hiroshi Esaki. LiVRation: Remote VR live platform with interactive 3D audio-visual service. In *Proceedings of the 2019 IEEE Games Entertainment & Media Conference*, pages 1–7, New York, NY, 2019. IEEE.
- Ryota Horie, Minami Wada, and Eri Watanabe. Participation in a virtual reality concert via brainwave and heartbeat. In *Proceedings of the 8th International Conference on Applied Human Factors and Ergonomics*, pages 276–284, Cham, Switzerland, 2018. Springer.
- Andrew MacQuarrie and Anthony Steed. Cinematic virtual reality: Evaluating the effect of display type on the viewing experience for panoramic video. In *Proceedings of the 2017 IEEE Conference on Virtual Reality*, pages 45–54, New York, NY, 2017. IEEE.
- Heather Bellini, Wei Chen, Masaru Sugiyama, Marcus Shin, Shateel Alam, and Daiki Takayama. Virtual & augmented reality: Understanding the race for the next computing platform. Goldman Sachs Global Investment Research, <https://www.goldmansachs.com/insights/pages/technology-driving-innovation-folder/virtual-and-augmented-reality/report.pdf>, January 2016.
- Angela Cresswell Jones, Rebecca Jane Bennett, and Samantha Cross. Keepin’ it real? Life, death, and holograms on the live music stage. In Angela Cresswell Jones and Rebecca Jane Bennett, editors, *The Digital Evolution of Live Music*, pages 123–138. Chandos Publishing, Oxford, UK, 2015.
- Alex Leavitt, Tara Knight, and Alex Yoshiba. Producing Hatsune Miku: Concerts, commercialization, and the politics of peer production. In Patrick W. Galbraith and Jason G. Karlin, editors, *Media*

- Convergence in Japan*, pages 200–229. Kinema Club, New Haven, CT, 2016.
- Ka Yan Lam. The Hatsune Miku phenomenon: More than a virtual J-pop diva. *The Journal of Popular Culture*, 49 (5): 1107–1124, 2016.
- Jan Stupacher, Pieter-Jan Maes, Matthias Witte, and Guilherme Wood. Music strengthens prosocial effects of interpersonal synchronization – if you move in time with the beat. *Journal of Experimental Social Psychology*, 72: 39–44, 2017.
- Ferda Ofli, Engin Erzin, Yücel Yemez, and Ahmet Murat Tekalp. Learn2Dance: Learning statistical music-to-dance mappings for choreography synthesis. *IEEE Transactions on Multimedia*, 14 (3-2): 747–759, 2012.
- Satoru Fukayama and Masataka Goto. Music content driven automated choreography with beat-wise motion connectivity constraints. In *Proceedings of the 12th International Conference in Sound and Music Computing*, pages 177–183, Genève, Switzerland, 2015. Zenodo.
- Taoran Tang, Jia Jia, and Hanyang Mao. Dance with melody: An LSTM-autoencoder approach to music-oriented dance synthesis. In *Proceedings of the 26th ACM International Conference on Multimedia*, pages 1598–1606, New York, NY, 2018. ACM.
- Daniel Thalmann and Soraia Raupp Musse. *Crowd Simulation*. Springer, London, UK, 2007.
- Mingliang Xu, Hao Jiang, Xiaogang Jin, and Zhigang Deng. Crowd simulation and its applications: Recent advances. *Journal of Computer Science and Technology*, 29 (5): 799–811, 2014.
- Masahiro Hamasaki and Masataka Goto. Songrium: A music browsing assistance service based on visualization of massive open collaboration within music content creation community. In *Proceedings of the 9th International Symposium on Open Collaboration*, pages 4:1–4:10, New York, NY, 2013. ACM.
- Masataka Goto. Music listening in the future: Augmented music-understanding interfaces and crowd music listening. In *Proceedings of the 42nd AES International Conference Semantic Audio*, pages 21–30, New York, NY, 2011. AES.
- Qunfang Wu, Yisi Sang, and Yun Huang. Danmaku: A new paradigm of social interaction via online videos. *ACM Transactions on Social Computing*, 2 (2): 7:1–7:24, 2019.
- Xiaojuan Ma and Nan Cao. Video-based evanescent, anonymous, asynchronous social interaction: Motivation and adaption to medium. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 770–782, New York, NY, 2017. ACM.
- Yue Chen, Qin Gao, and Pei-Luen Patrick Rau. Watching a movie alone yet together: Understanding reasons for watching Danmaku videos. *International Journal on Human-Computer Interaction*, 33 (9): 731–743, 2017.
- Daniel Johnson. Polyphonic/pseudo-synchronic: Animated writing in the comment feed of Nicovideo. *Japanese Studies*, 33 (3): 297–313, 2013.
- Masataka Goto. A chorus section detection method for musical audio signals and its application to a music listening station. *IEEE Transactions on Audio, Speech & Language Processing*, 14 (5): 1783–1794, 2006.
- Ivan Laptev, Marcin Marszalek, Cordelia Schmid, and Benjamin Rozenfeld. Learning realistic human actions from movies. In *Proceedings of the 2008 IEEE Conference on Computer Vision and*

- Pattern Recognition*, pages 1–8, New York, NY, 2008. IEEE.
- François Petitjean, Alain Ketterlin, and Pierre Gançarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44 (3): 678–693, 2011.
- Masaharu Mori, Tomoaki Moriya, and Tokiichiro Takahashi. e-Ovation: A live karaoke system by using virtual reality tools. In *Proceedings of the 2017 International Workshop on Advanced Image Technology*, pages 3A3:1–3A3:4, Selangor, Malaysia, 2017. Multimedia University Press.
- Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A highly efficient gradient boosting decision tree. In *Proceedings of the 31st Annual Conference on Neural Information Processing Systems*, pages 3146–3154, San Diego, CA, 2017. NeurIPS Foundation.
- George Tzanetakis and Perry R. Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10 (5): 293–302, 2002.
- Masataka Goto, Kazuyoshi Yoshii, Hiromasa Fujihara, Matthias Mauch, and Tomoyasu Nakano. Songle: A web service for active music listening improved by user contributions. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, pages 311–316, Montreal, Canada, 2011. ISMIR.
- Mel Slater, Amela Sadagic, Martin Usoh, and Ralph Schroeder. Small-group behavior in a virtual and real environment: A comparative study. *Presence: Teleoperators and Virtual Environments*, 9 (1): 37–51, 2000.
- Frank Biocca, Chad Harms, and Jennifer Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *Proceedings of the 4th Annual International Workshop on Presence*, pages 1–9, Philadelphia, PA, 2001. ISPR.
- Kristine L. Nowak and Frank Biocca. The effect of the agency and anthropomorphism on users’ sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 12 (5): 481–494, 2003.
- YoungChan Hwang and Joon Soo Lim. The impact of engagement motives for social TV on social presence and sports channel commitment. *Telematics and Informatics*, 32 (4): 755–765, 2015.
- Jaclyn Ann Cameron. *The On-Screen Water Cooler: Effects of Televised User-Generated Comments on Cognitive Processing, Social Presence, and Viewing Experience*. PhD thesis, University of Tennessee, Knoxville, TN, August 2016.
- Li Chen. How danmaku influences emotional responses: Exploring the effects of co-viewing and copresence. Master’s thesis, Nanyang Technological University, Nanyang Avenue, Singapore, February 2018.
- Linjia He, Hongsong Li, Tong Xue, Deyuan Sun, Shoulun Zhu, and Gangyi Ding. Am I in the theater?: Usability study of live performance based virtual reality. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pages 28:1–28:11, New York, NY, 2018. ACM.
- Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. Using presence questionnaires in virtual reality. In *Proceedings of the 2019 ACM CHI Conference on Human Factors in Computing Systems*, pages 360:1–360:12, New York, NY, 2019. ACM.

- Kit Yan To. The voice of the future: Seeking freedom of expression through VOCALOID fandom. Master's thesis, The University of Texas at Austin, Austin, TX, May 2014.
- Lucas Maystre and Matthias Grossglauser. Fast and accurate inference of Plackett-Luce models. In *Proceedings of the 29th Annual Conference on Neural Information Processing Systems*, pages 172–180, San Diego, CA, 2015. NeurIPS Foundation.
- James Durbin. Incomplete blocks in ranking experiments. *British Journal of Statistical Psychology*, 4 (2): 85–90, 1951.
- William Jay Conover. *Practical Nonparametric Statistics*. John Wiley and Sons, New York, NY, 3 edition, 1999.
- Meinard Müller, Daniel P. W. Ellis, Anssi Klapuri, and Gaël Richard. Signal processing for music analysis. *IEEE Journal of Selected Topics in Signal Processing*, 5 (6): 1088–1110, 2011.
- Robert B. Zajonc. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9 (2): 1–27, 1968.
- Robert B. Zajonc. Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science*, 10 (6): 224–228, 2001.