

歌詞没入型身体的音楽体験を可能とする 光学透過式MR環境LyricBathe

小島 颯英^{1,a)} 中小路 久美代¹ 加藤 淳² 後藤 真孝²

概要：LyricBathe は、HoloLens 2 を用いて歌詞を楽曲と同期して空間に表示するインタラクティブな MR 環境である。24 種の異なるモードによる歌詞体験の違いを、注視箇所と頭部位置データをもとに調査した。予備実験と本実験を通して、モードによって視線行動が一貫して誘発される場合と、個人差が表れやすい場合があることが明らかになった。また、曲調とモードの組み合わせによって体験スタイルが変化する傾向も確認された。これらの知見は、ユーザ特性に応じた MR 歌詞体験の最適化や、より没入感の高い体験設計に向けた基盤的知見を提供する。

1. はじめに

歌詞は音楽を豊かにする詩的表現であり、メディア技術の発展とともに歌詞カード、音楽再生アプリケーションのカラオケ的な歌詞表示、より多様なアニメーション（キネティックタイポグラフィ）を含むリリックビデオ、インタラクティブな「リリックアプリ (lyric app)」のように様々な視覚表現として楽しまれてきた。加藤ら [1] は、リリックアプリを提唱するとともに、その開発を支援する Lyric App Framework を提供し、楽曲の再生と同期した多様な歌詞体験を実現するアプリケーション開発を支援してきた。しかし、既存のリリックアプリの多くは、パーソナルコンピュータやスマートフォンなどの一般的な矩形ディスプレイを前提としている。

一方、本研究では、歌詞を身体的に体験する没入型の新たな音楽鑑賞スタイルを可能にすることを目指し、ユーザが光学透過型ヘッドマウントディスプレイ（HoloLens 2）を用いた複合現実（MR）環境内で歌詞を楽しむシステム「LyricBathe」を開発してきた。本システムでは、歌詞は楽曲の進行と同期する 3DCG オブジェクトとして表示され、ユーザの身体や現実世界の物理的なオブジェクトとインタラクションすることが可能である。例えば、文字が体から飛び出してきたり、文字を浴びられるように落ちてきたり、ユーザの周りを囲むように文字が表示されたりするような身体的なインタラクションを実現するモードを用意しており、歌詞に「没入」ような歌詞体験が可能である

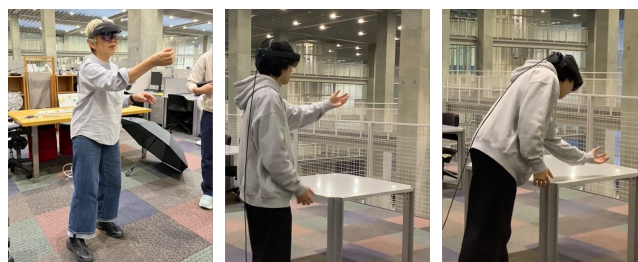


図 1 LyricBathe 体験中のユーザの様子

(図 1)。

これまでに、歌詞の文字の大きさ、動き、表示タイミング、身体や現実世界とのインタラクションの方法などの表現方法（インタラクションプリミティブ）を様々な組み合わせた種々のモードを実装してきた。当初は、プロジェクトの初期段階で構築した 8 つのモードについて注視箇所を計測するユーザ調査を実施した。しかし、いずれのモードが魅力的だと評価されているか、有意な結論を導くことはできなかった [2]。そこで、新たなモードの開発を継続し計 24 種類のモードを実装し、歌詞没入型身体的音楽体験に寄与する基本的な要素をインタラクションプリミティブとして 11 項目、定性的な議論をもとに同定した [3]。

本論文では、各インタラクションプリミティブのユーザ体験への寄与や、楽曲の特徴による影響について調査するために、2 つのユーザ観察実験を行った。まず、1 曲の歌詞付き楽曲を用いて、1 名のユーザを対象に、LyricBathe の全 24 種類のモードについて、注視行動や頭部運動といったユーザの行動を記録・分析する予備実験を行った。具体的には、実験で記録したデータについてモードごとのクラスタリングを行い、歌詞体験を構成するインタラクション

¹ 公立はこだて未来大学

² 産業技術総合研究所

^{a)} g2124009@fun.ac.jp

プリミティブによるユーザ体験の変化を分析した。予備実験では、巨大な文字を表示するモードではユーザの頭部運動が大きくなるなど、インタラクションプリミティブが異なる行動特徴を誘起することが示唆された。続いて、予備実験で特徴的な行動が見られた8種のモードにおいて、同一の歌詞を使用したロック調、ヒップホップ調、バラード調といった3種類の楽曲を用いて、4名のユーザを対象に本実験を実施した。本実験を通して、参加者や楽曲によって異なる行動の特徴がみられるモードが存在すること、定性的なインタビュー結果からは、静的な集中から動的な探索まで、参加者ごとに様々な評価軸が存在し、注視パターンにそれらが表れることが明らかとなった。

2. 関連研究

さまざまなテキスト表現形式の中でも、歌詞の表現は、音楽の時間的進行によって制約を受ける形式として際立っている。テキストの動きを同期させたり、歌詞の特定の部分を強調したりすることで、音楽鑑賞体験をより豊かなものにすることができる。本章では、MR環境におけるテキストのダイナミックな表現に関連する既存研究について概説する。音楽に付随する空間的な表現やインタラクティブな表現はさまざまな形で研究されてきたが、MR環境でインタラクティブかつ身体的な体験を創出する研究は少なく、そうした表現を設計するための知見にも乏しい。

2.1 キネティックタイポグラフィ

テキストに、動き、回転、変形を加えてアニメーションさせる視覚表現は、キネティックタイポグラフィと総称される。Shannonら[4]は、1990年代後半にはすでに、そのデザインの基礎となるいくつかの原則について論じている。30年以上にわたる経験と知識の蓄積により、キネティックタイポグラフィは、歌詞が楽曲再生にあわせてタイミングよくアニメーションする動画（リリックビデオ）など、テキストのダイナミックな視覚表現を創り出す技術として広く利用されるようになった。キネティックタイポグラフィは主に二次元の動画として書き出される静的な視覚表現であり、ユーザ自身が自由に動き回れるような三次元のMR環境にそのまま適用することはできない。

2.2 楽曲再生と同期した歌詞表現

楽曲再生と同期した歌詞テキストの表示は、カラオケやリリックビデオでよく見られる。リリックアプリは、こうした表現にユーザとのインタラクションを加えた新たな歌詞駆動型の視覚表現である[1]。例えば、特定の楽曲の世界観を表現したVR空間を体験できる拡張現実アプリ^{*1}や、

ユーザが画面上の歌詞を左右にドラッグ&ドロップして楽曲の再生位置をコントロールできるインタラクティブなリリックビデオ^{*2}などがある。また、身体性を伴う体験を提供するリリックアプリとして、スマートフォンのカメラで撮影した自撮り映像上で、口からリアルタイムに再生中の歌詞が湧き出てくるもの^{*3}がある。ただし、これらの例があくまでパーソナルコンピュータやスマートフォンなどの平面的な矩形ディスプレイ上に表示されるものであるのに対し、本論文で取り組んでいる光学透過型ヘッドマウントディスプレイを使用する複合現実（MR）環境では、歌詞が直接知覚される現実世界の空間に重ねて表示されるため、考慮すべき要素が異なる。例えば、歌詞がユーザの身体や、家具、手持ちの物体、他の人物など周囲の物理的環境と相互作用することが考えられ、本研究ではこうした従来のキネティックタイポグラフィを中心とした歌詞表現を超える体験設計に有用な知見の創出を目指している。

2.3 MR環境でのテキスト体験

文字をMR空間内で体験する既存研究は多数ある。例えば、物理的なオブジェクトと仮想のテキスト表示を組み合わせた文字表現として、横濱ら[5]の構築したMAVenReadがある。MAVenReadでは、HoloLens 2を装着して物理的な本を開くと、ほぼ白紙のページ端に印刷されたARマークが読み取られ、そのページ上に詩の文字が動的に重畳表示される。重畳表現される文字を一定方向に動かしたり、ゆっくりと回転させるなどのアニメーションにより、新たな読詩体験を実現している。また、VRやMRを用いて複数の文書を空間的に配置し、文書の潜在的な意味を抽出するような取り組みがある。Bandyopadhyay[6]は、VRヘッドマウントディスプレイを用いた文書の意味抽出のためのシステムとしてImmersive Space to Think (IST)というシステムを構築し、没入型3D空間での文書の意味抽出プロセスの有用性について検証を行っている。Weizhouら[7]は、オフィス環境でのMR技術を用いた文書の空間配置について、ユーザ中心デザイン研究を実施し、インタラクション手法を抽出している。しかしながら、これら既存のMR環境における文字表現の手法は、動的に変わる音楽の進行に同期する形での歌詞表現と比べると、躍動感や時間厳密性といった点において、ユーザの体験が本質的に異なると考えられる。

2.4 モーショングラフィックス

モーショングラフィックスはデータや画像をアニメー

^{*1} うたろ: キミを探す、夏 (初音ミク「マジカルミライ 2021」プログラミング・コンテスト).
<https://textalive.jp/apps/Ew3oA50F9GA4YyMF>

^{*2} catLee: Lyric Trail (初音ミク「マジカルミライ 2020」プログラミング・コンテスト).

<https://textalive.jp/apps/hIuZ6YCsV0If1gBo>

^{*3} zabu: lyric-from-mouth (初音ミク「マジカルミライ 2023」プログラミング・コンテスト).

<https://textalive.jp/apps/OHQeAKG7bIJ3iB8e>

ションで表示する映像ベースの情報表現手法の一つである。Amir ら [8] は、モーショングラフィックスを組み込んだ映像制作のためのオーサリングツールを配布し、モードがデザインプロセスに与える影響について調査した。プロのモーションデザイナーはカスタマイズできるオプションが多いことを好むなど、専門性の高さによってモードに求める要件が異なることを示すものだった。モーショングラフィックスは、あらかじめ映像をすべてデザインして制作する必要があり、動的に体験を創出する本研究とは体験のデザインが異なるものとなる。

2.5 ジェネレイティブアート

ジェネレイティブアートは、数学的アルゴリズムとコンピュータプログラムを使用してデザインを生成するアートの一種である。David らは、ジェスチャと音声を使って複数人が同じジェネレイティブアートとインタラクションできるワイヤレスポータブルデバイス i-Cube を構築している [9]。i-Cube を通してユーザの物理的な動きや音声を読み取ることで、インタラクティブかつダイナミックな視覚表現をスクリーンに表示することが出来る。これに対して LyricBathe は、一人で音楽鑑賞を行うことを前提に構築しており、表示空間に関してもスクリーンのような平面ではなく、3次元空間に歌詞の表示を行うことが特徴的である。

2.6 空間的な音楽体験の拡張

音楽の物理空間内での表現に関しては、5.1ch のサラウンドや空間オーディオなどが一般化している。音楽を物理空間内でインタラクティブに楽しむ研究として Miyagawa ら [10] は、物理空間に音源を自在に配置して音楽鑑賞を行うシステムを構築している。この研究では、MR 環境でオーケストラの音の分布を空間にビジュアルとして表現し、各音源を空間に固定したり、ユーザが自由に配置しながらインタラクティブに音楽を楽しむ音楽体験を実現している。Jongik ら [11][12] の構築した SoundMist では、音を噴霧する音楽体験を実現している。物理的なスプレーを空間に噴霧することで、空間内に音源を配置し、霧と共に霧散していくような没入感のある空間的・聴覚的体験を作り出している。しかし、これらの空間的な音楽体験を拡張する既存研究は、あくまで音の体験を拡張するものであり、本研究で目指している楽曲本来の音楽的要素を損なわないような音楽体験とは異なると考えられる。

3. LyricBathe

LyricBathe は、光学透過式の MR デバイスである HoloLens 2 を使用した MR システムであり、歌詞を 3DCG オブジェクトとして MR 環境内に、音楽の進行と同期しながら表示する。それに加えて、ユーザが身体を動かすことで歌詞とインタラクションすることができる。ユーザ

は、MP3 形式の楽曲と歌詞のテキストデータがあれば、LyricBathe の各モードを体験することができる。

歌詞と音楽を同期させるために、LyricBathe はリリックビデオ・アプリの制作・配信支援プラットフォームである TextAlive [13] と共通の歌詞タイミング解析技術を利用している。これにより、各文字の発音タイミング、各文字のフレーズ ID、各フレーズの終了タイミングをリスト化した CSV ファイルが生成される。この CSV ファイルは、Unity 上で動作する LyricBathe によって読み込まれ、同時に音楽の音声ファイルも再生される。LyricBathe は、歌詞表現の最も原始的な要素、すなわち歌詞の各フレーズと、各文字の発音の開始と終了のタイミングに焦点を当てている。歌詞のタイミング、品詞、音楽的な印象、ビート構造等からリリックアプリを構築することが出来る Lyric App Framework と比較すると、LyricBathe ははるかにシンプルなメカニズムを採用している。

現在、LyricBathe には 24 のモードがあり、それぞれ異なるインタラクティブモードを備えている。身体とのインタラクション、文字の大きさ、動き方などの側面に焦点を当て、24 種のモードの実装を通して、LyricBathe における歌詞表現のデザイン空間を探索している (表 1)。

4. 予備実験

LyricBathe のモードを用いてインタラクションプリミティブの音楽体験に与える影響について調査し、本実験で用いるモードを選定するため予備的ユーザ観察実験を行った。

4.1 予備実験で使用した楽曲

予備実験を行うにあたって Suno AI [14] の歌詞生成機能を用いて歌詞を生成し、Suno AI によりその歌詞付きの楽曲を生成した。実験では、生成された楽曲から 1 コーラス分のみを切り取り、曲の長さが 1 分 18 秒になるよう編集したものを利用した。

実験結果として記録したデータでは、試行ごとに実験データの終了時刻にばらつきが見られたため、再生開始から 1 分 15 秒の範囲に調整して分析を行った。

4.2 実験条件

大学内で参加者を募集し、20 代大学院生男性 1 名を対象に実験を行った。実験環境としては、ユーザの正面方向に正方形の机を一つ配置し、それ以外の障害物を配置しない空間で実験を行った。

4.3 実験におけるデータの計測

予備実験では、体験中のユーザの行動に関する 2 種類のデータを、それぞれ CSV ファイルとして記録した。

- ユーザの注視箇所に関する時系列データ

表 1 LyricBathe で実装した 24 種類のモードとそれを構成するインタラクションプリミティブ

Interaction Primitives	1	2	3	4	5	6	7	8	9	10	11
	How to Arrange Characters	Where to Display Characters	Scaling Characters	When to Display Characters	How Long to Keep Displaying Characters	Units for Defining Behavior	Character Movement	Phrase Movement	User Interaction	Whether virtual objects have physical properties	When to Apply Physical Properties
modes											
M01	straight line	relative positioning	1x	character-based	disappeared at the end of each phrase	phrase	n/a	flowing out from body	finger-based gestures	No	n/a
M02	straight line	relative positioning	1x	character-based	disappeared at the end of each phrase	phrase	n/a	flowing between fingers	finger-based gestures	No	n/a
M03	straight line	absolute positioning	1x	character-based	25 seconds after the end of the phrase	character	falling	n/a	palm-based gestures	Yes	at the end of each phrase
M04	straight line	absolute positioning	1x	character-based	25 seconds after the end of the phrase	character	floating	n/a	palm-based gestures	Yes	at the end of each phrase
M05	straight line	absolute positioning	1x	character-based	25 seconds after the end of the phrase	character	directing	n/a	palm-based gestures	Yes	at the end of each phrase
M06	straight line	absolute positioning	1x	character-based	25 seconds after the end of the phrase	character	Throwing away	n/a	finger-based gestures palm-based gestures	Yes	at the end of each phrase
M07	straight line	absolute positioning	1x	song-based	25 seconds after the end of the phrase	character	gathering to phrase	n/a	palm-based gestures	Yes	when each character appears
M08	straight line	relative positioning	1x	character-based	25 seconds after the end of the phrase	character	shower	n/a	palm-based gestures	Yes	at the end of each phrase
M09	straight line	absolute positioning	5x	character-based	25 seconds after the end of the phrase	character	falling	n/a	palm-based gestures	Yes	at the end of each phrase
M10	straight line	absolute positioning	15x	character-based	25 seconds after the end of the phrase	character	falling	n/a	palm-based gestures	Yes	at the end of each phrase
M11	straight line	absolute positioning	1x	song-based	25 seconds after the end of the phrase	character	gathering to phrase	n/a	palm-based gestures	Yes	when each character appears
M12	straight line	relative positioning	1x	character-based	25 seconds after the end of the phrase	character	falling	n/a	palm-based gestures	Yes	when each character appears
M13	straight line	absolute positioning	1x	character-based	25 seconds after the end of the phrase	phrase	n/a	falling	palm-based gestures	Yes	at the end of each phrase
M14	straight line	absolute positioning	1x	character-based	25 seconds after the end of the phrase	phrase	n/a	floating	palm-based gestures	Yes	at the end of each phrase
M15	straight line	absolute positioning	1x	character-based	25 seconds after the end of the phrase	phrase	n/a	directing	palm-based gestures	Yes	at the end of each phrase
M16	straight line	absolute positioning	1x	character-based	25 seconds after the end of the phrase	phrase	n/a	throwing away	finger-based gestures palm-based gestures	Yes	at the end of each phrase
M17	straight line	absolute positioning	1x	phrase-based	25 seconds after the end of the phrase	character	falling	n/a	palm-based gestures	Yes	at the end of each phrase
M18	straight line	absolute positioning	1x	phrase-based	25 seconds after the end of the phrase	character	floating	n/a	palm-based gestures	Yes	at the end of each phrase
M19	straight line	absolute positioning	1x	phrase-based	25 seconds after the end of the phrase	character	directing	n/a	palm-based gestures	Yes	at the end of each phrase
M20	straight line	absolute positioning	1x	phrase-based	25 seconds after the end of the phrase	character	Throwing away	n/a	finger-based gestures palm-based gestures	Yes	at the end of each phrase
M21	circular shape	absolute positioning	1x	character-based	25 seconds after the end of the phrase	character	falling	n/a	palm-based gestures	Yes	at the end of each phrase
M22	random	absolute positioning	random scaling	character-based	25 seconds after the end of the phrase	character	falling	n/a	palm-based gestures	Yes	at the end of each phrase
M23	finger tracking	absolute positioning	1x	character-based	25 seconds after the end of the phrase	character	falling	n/a	finger-based gestures palm-based gestures	Yes	at the end of each phrase
M24	finger tracking	absolute positioning relative positioning	1x	character-based	25 seconds after the end of the phrase	character	falling	n/a	finger-based gestures palm-based gestures	Yes	at the end of each phrase

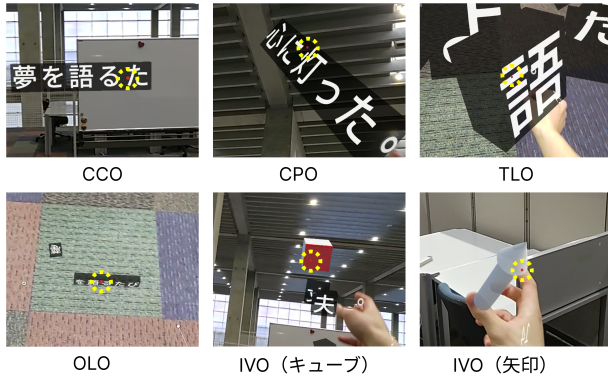


図 2 注視箇所タグごとの代表例

● ユーザの頭部位置に関する時系列データ

注視箇所のデータは、HoloLens 2 の Eye Tracking API により取得された視線情報をもとに、注視対象のタグ（詳細は次節で説明）と注視時間を CSV 形式で記録した。頭部位置のデータは、楽曲の再生開始から 0.2 秒ごとに、MR 空間内におけるユーザが装着しているデバイスの 3 次元座標（x: 左右方向, y: 上下方向, z: 奥行き方向）を取得し、同様に CSV 形式で記録した。

4.4 注視箇所の記録方法

予備実験では、Eye Tracking API により取得された視線データから、ユーザが注視している歌詞オブジェクトやその他の対象に対して、以下に示す注視タグを割り当てた。各タグをもとに、注視箇所の時系列遷移を記録し、時間の経過やインタラクションに応じて変化する注視対象の推移をデータとして取得した（図 2 参照）。

- CCO (Current Character Object): その時点で発声されている歌詞の文字上または表示から 1 秒以内の歌詞オブジェクト上に視線がある
- CPO (Current Phrase Object): その時点で発声されている歌詞の文字が入ったフレーズ上に視線がある (CCO 以外)
- TLO (Touched Lyric Object): 手指で操作している歌詞の上に視線がある
- OLO (Other Lyric Object): それ以外の歌詞の文字の上に視線がある (CCO でも CPO でも TLO でもないもの)
- IVO (Interaction Virtual Object): 表示されているユーザインタフェースオブジェクト（シャワーキューブや矢印オブジェクト）の上に視線がある
- UNO (Undefined Natural Observation): どの状態にも当てはまらない場所を注視している

CSV ファイルには、注視箇所タグが切り替わるごとに注視開始と注視終了のタイムスタンプ、注視箇所のタグをそれぞれ記録した。タイムスタンプの形式は mm:ss.ms の

形式で文字列として記録した。

4.5 予備実験の実施

予備実験では、インタラクションプリミティブによる歌詞体験の影響を調査するため LyricBathe のモード群の全 24 モードを対象として実験を行った（表 1）。この実験では、HoloLens 2 を装着してもらい、立った状態から LyricBathe の各モードを体験してもらった。体験中の参加者の動きに制限はかけずに実験を行った。

24 種類のモードは、ランダムに並び変えて体験してもらった。モードの切り替え操作は、実験者がワイヤレスキーボードを用いて遠隔で行った。各モードで行えるインタラクションについては特に説明をせず、遠隔で楽曲の再生を開始した。

4.6 予備実験結果に対する注視行動に基づくユーザ行動分析

歌詞やインタラクションに対する注視行動について分析するために視線タグの UNO は視線データから除外し、「CCO、CPO、OLO、TLO、IVO」だけの注視箇所の遷移について可視化を行った。

注視箇所の遷移傾向について K-means クラスタリングを行った結果、24 種のモードは、エルボー法により以下の 4 つのクラスタに分類された。

クラスタ 0: M11,M18,M17,M16,M14,M13,M23,M20,M10,M21,M07,M06,M04,M03,M22,M09

このクラスタのモードでは、CCO と CPO 間の遷移を基本としつつ、OLO を中心とした遷移も見られた。この結果は、現在のフレーズに注目しつつも時折周囲を確認し、歌詞を楽しむ体験が誘発されていると考えられた。

クラスタ 1: M05,M15,M19

このクラスタのモードでは、IVO からその他の注視タグ (CCO、CPO、OLO) への遷移が多く含まれていた。この結果は、矢印型の UI オブジェクトを注視しながら操作し、歌詞オブジェクトの変化を楽しむ体験が誘発されていると考えられた。

クラスタ 2: M08,M12

このクラスタのモードでは、CCO から TLO への遷移が含まれていた。この結果より、音楽と共に表示された歌詞を手で触れるようなインタラクションを誘発されていると考えられた。

クラスタ 3: M01,M02,M24

このクラスタのモードでは、CCO と CPO 間の遷移がほとんどを占めていた。この結果は、フレーズの細部と全体を行き来する視線行動が誘発されていると考えられた。

4.7 予備実験結果に対する頭部位置に基づくユーザ行動分析

頭部位置の評価指標として、それぞれの軸についての標準偏差と座標の範囲に着目した。頭部位置の特徴量についての K-means クラスタリングを行った結果、24 種のモードは、エルボー法により 5 つのクラスタに分類された。

クラスタ A: M01,M15,M08,M05,M12,M02

このクラスタのモードは x 座標と z 座標の標準偏差と値の範囲が小さくなっていた。これは、あまり移動を行わず、静的に音楽鑑賞を行っていることを示していると考えられた。

クラスタ B: M20,M19,M18,M17,M16,M04,M14,M06,M23,M11,M03,M24

このクラスタのモードは、x 座標の標準偏差と範囲が小さく、z 座標の標準偏差と範囲が大きくなっていた。これは、前後の移動を多く行いながら音楽鑑賞を行っていることを示していると考えられた。

クラスタ C: M10

このクラスタでは、他のモードと比べて x 座標と z 座標の標準偏差と範囲が極端に大きくなっていた。これは、大幅に移動しながら音楽鑑賞を行っていることを示していると考えられた。

クラスタ D: M07,M21,M13

このクラスタに含まれるモードは、y 座標の標準偏差と範囲が大きくなっていた。これは上下の移動が多く行われていたことを示し、床に落ちた文字にインタラクションを行っていることを示していると考えられた。

クラスタ E: M09,M22

このクラスタのモードはクラスタ C ほどではないが、x 座標と z 座標の標準偏差と範囲が大きくなっていた。これは、移動しながら音楽鑑賞を行っていることを示していると考えられた。

5. ユーザ観察本実験

予備実験の分析結果を踏まえて、ユーザ行動の特徴を表すクラスタを代表するモードを 8 種類選定した (図 3)。モード、楽曲のジャンル、参加者の 3 つの変数による歌詞体験の変化を観察するため、ユーザ観察本実験を行った。本実験では、「ある曲に適したモードがある」「モードによって誘発される体験が決まる」という仮説に基づき、LyricBathe における歌詞体験の構造を明らかにすることを目的として実施した。

5.1 本実験で使用した楽曲

本実験を行うにあたって、Suno AI [14] を用いて日本語の歌詞の楽曲を生成した。楽曲の歌詞は、「前を向くための原動力として過去を振り返る」というプロンプトで、18 フレーズからなる歌詞を ChatGPT を用いて生成したも

のを利用した。Suno AI を用いて同一の歌詞を用いてジャンルの異なる 3 つの楽曲を生成した。生成する楽曲ジャンルとして楽曲のテンポ感に着目し、バラード (BPM: 90)、ヒップホップ (BPM: 110)、ロック (BPM: 160) の 3 種類のジャンルの楽曲を生成した。実験では、生成した楽曲から 1 コーラス分を切り取り、曲の長さが 47 秒になるように編集したものを使用した。分析にあたっては、実験で記録したデータのうち再生開始から 45 秒の範囲を使用した。

5.2 実験手順

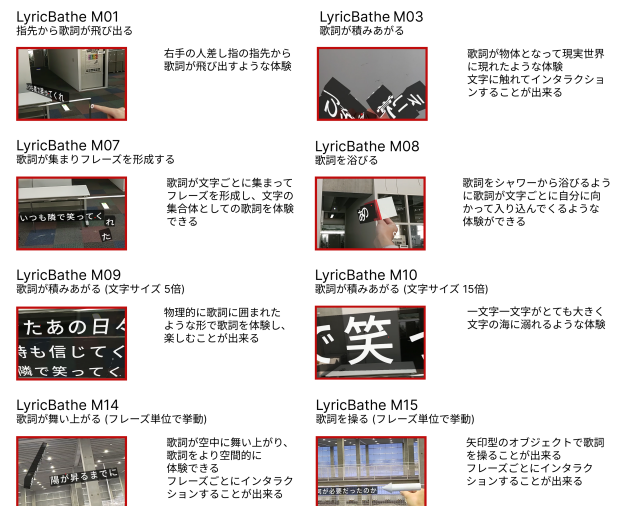


図 3 本実験で用いた LyricBathe の 8 種のモード

実験参加者は、大学生の男性 3 名と、大学教員の女性 1 名の計 4 名である。実験環境として日常生活における壁や机などの周囲環境を想定し、左右に仕切りと机がある環境で実験を行った。実験環境に基準点を設定し、実験では参加者に 1 試行ごとに基準となる位置に戻ってもらい、LyricBathe を体験してもらった。

本実験では、M01、M03、M07、M08、M09、M10、M14、M15 の 8 モード (図 3) において、歌詞が同一でジャンルの異なる 3 楽曲を用いたユーザ観察実験を行った。参加者には HoloLens 2 を装着してもらい、視線追跡のキャリブレーションを行った後、基準点に立った状態で LyricBathe を起動してもらい、各モードを体験してもらった。

モードの体験順は全参加者で統一し、M01、M03、M07、M08、M09、M10、M14、M15 の順で実施した。各モードにおける 3 楽曲の体験順は参加者ごとにランダムに設定し、計 24 試行 (8 モード × 3 楽曲) を行った。モードおよび楽曲の切り替えは、実験者がワイヤレスキーボードを用いて遠隔で操作した。モードの切り替えごとに、そのモードで可能なインタラクションについて簡単な説明を行い、その後楽曲を再生した。

また、予備実験と同様に、ユーザのしている映像データ、

体験中の注視箇所と頭部位置に関する 2 つの CSV ファイルを記録した。

各参加者には、LyricBathe の体験終了後、以下の質問を含む半構造化インタビューを実施した。

- 印象に残ったモードはどのモードか
- いまいちだと感じたモードがあったか
- モードによって取りたくなった行動があるか
- 楽曲ごとにマッチしたモードがあったか

6. 本実験の結果と分析

本節では、実験データの分析結果を示す。なお、以下では各要素を以下のように表記する：Mxx はモードの ID (M01、M03、M07、M08、M09、M10、M14、M15)、Sxx は楽曲の ID (S01：ロック調、S02：ヒップホップ調、S03：バラード調)、Pxx は参加者の ID (P01、P02、P03、P04) を表す。

6.1 インタビュー結果

各参加者のインタビュー結果を表 2 に示す。

表 2 各参加者へのインタビュー結果

参加者	発言
P01	しんみりした曲は歌詞のオブジェクトを裏側から見たりしてじわっと来た
P01	「約束」とかを守りたくなる
P01	床に残っている歌詞がウツとくる
P01	曲の違いはスピード、速い曲をシャワーで体験するとマシニングのようだった
P01	矢印のやつ (M15) は離れたところからよりも文字のところからやりたくなった
P02	揺らめいている (M14) のが歌詞に浸かっている感じが一番よかった
P02	特にあまり激しい曲調じゃないものが一番好み
P02	下に落ちているのにはあまり興味をそられなかった
P02	落ちている歌詞よりも今流れている歌詞に集中したい
P02	浮いているやつが良い目で追いやすい
P03	面白かったのは飛んでいく方向を矢印で決めるやつが思ったより飛んで行って面白かった
P03	ロックっぽいやつ (S01) が良かった 理由：感覚的に一番頭に残った
P03	歌詞をつまみだして面白くて座り込んで触ったりしていた
P03	(見えていないと思っても) 視点を変えると文字があるということがあった
P04	一番でかくでくるやつ (M10) が一番良かった
P04	視覚的にインパクトがあるのど下に溜まっていくというのが面白い
P04	溜まったやつをどかしたりできるのが面白かった
P04	ゆっくりな方が見てて触って一番心地よかったのでバラード (S03) が良かった
P04	矢印のやつ (M15) はどの曲でも合う

6.2 注視割合による体験スタイルの多様性の抽出

LyricBathe における視線行動の特徴を把握するため、まず注視箇所 (ラベル：CCO、CPO、OLO、TLO、IVO) ごとの注視時間割合を分析した。この分析の目的は、同じモードでも参加者ごとに視線行動が異なるのか、あるいはモードが視線行動を一様に誘発するのかを明らかにすることである。

具体的には、(参加者 ID × モード ID × 楽曲 ID) ごとの注視割合ベクトルを抽出し、モードと楽曲の組み合わせごとに 4 名の参加者間でのばらつきを計算した。ばらつきの指標としては、各注視ラベルの割合の標準偏差を用いた。

図 4 には、モードと楽曲ごとの参加者別の注視箇所の割合を棒グラフで表したものである。ここからは、M01 で

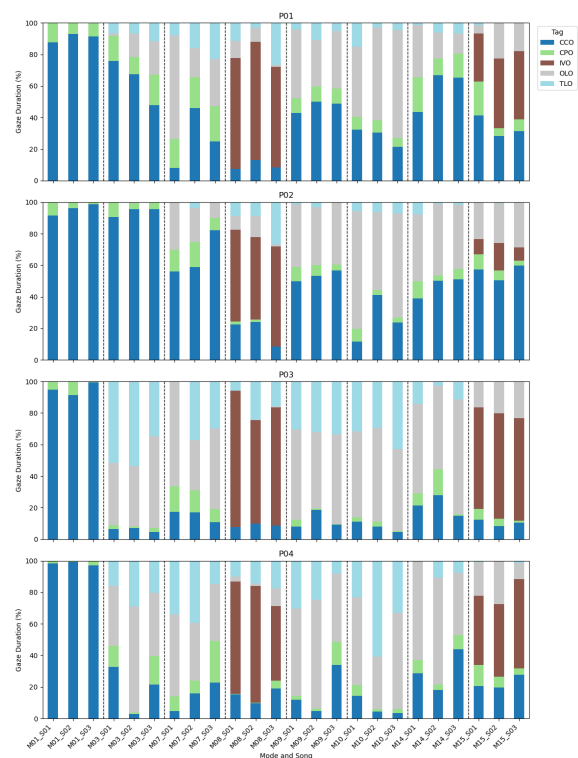


図 4 参加者ごとのモードと楽曲の組み合わせにおける注視タグごとの注視割合

は、CCO/CPO への注視割合が高く、全員の注視が発声中の歌詞に集中していることがわかる。M08 は全参加者で IVO の割合が多く、ユーザインタフェースオブジェクトへ注視が集中していることがわかる。一方で、M03 や M09 では注視割合の構成が参加者によって異なり、OLO や TLO への注視が多い参加者もいれば、CCO/CPO への集中が見られる参加者も見られた。

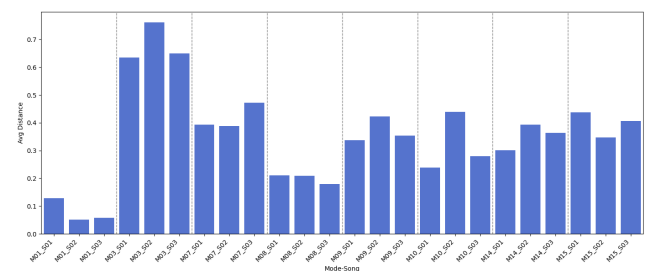


図 5 各モードにおける注視割合の参加者間の違いに関するユークリッド距離

これらの差異を定量的に示したのが図 5 である。図 5 では、縦軸の値として各モードにおける注視割合の参加者間の違いに関するユークリッド距離を示している。この結果から、M01、M08 のようなモードでは参加者間の差が小さく、視線行動が設計によって強く統制されていることがわかる。一方、M03、M07 などは高いばらつきを示し、ユーザの選好に応じた視線行動が現れていることが見て取れた。この分析により、モードによって視線行動の自由度に違

いがあり、それが体験の多様性に直結していることが示された。

6.3 頭部位置データに基づく空間探索行動の分析

視線に加え、身体の動きによる空間的探索行動を明らかにするため、各試行における頭部位置データを分析した。目的は、ユーザがどの程度空間的に移動しながら体験していたかを定量化し、視線と同様にモードや楽曲の影響が現れるかを検証することである。分析方法としては、x, y, z 座標の標準偏差と範囲（最大－最小）をそれぞれ算出し、以下の式によって「空間探索スコア」を定義した。

空間探索スコア =
$$\frac{x_{range} + y_{range} + z_{range}}{x_{std} + y_{std} + z_{std}} \quad (1)$$

このスコアは、ユーザがその場で揺れるように動いていた場合は低くなり、逆に目的をもって移動していた場合は高くなると解釈できる。

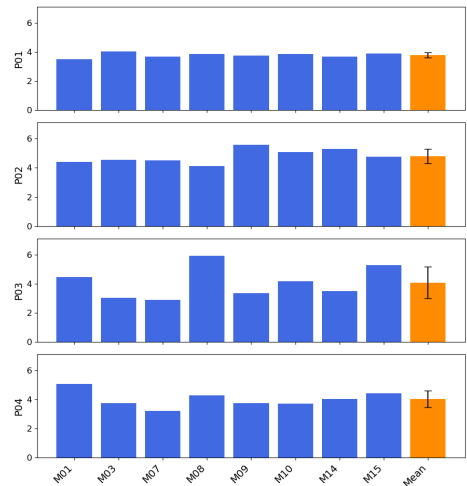


図 6 参加者ごとの各モードの探索スコア（楽曲平均）

図 6 は、各参加者のモードごとの空間探索スコアをプロットしたものである。図から、P01 は全体的にスコアが低く、あまり移動していなかったことが読み取れる。一方、P02 は全体的にスコアが高く、特に M09～M14 において積極的に空間を移動していたことが読み取れる。P03 や P04 は一部のモードでスコアが高かったが、モードごとのばらつきが見られた。これらの結果から、空間探索行動はモードや楽曲に応じた変化よりも、個人差の方が顕著であることが確認された。

表 3 の参加者間のモード×楽曲ごとの 6 次元ベクトル（各軸の標準偏差と範囲）のユークリッド距離を見ても、モード間や楽曲間を横断する傾向は見られなかった。このことから、頭部位置データは視線ほどモードや楽曲の設計効果を反映せず、個人の探索スタイルや姿勢のクセが強く現れることが示唆された。すなわち、視線は「モードに応じた体験」を捉える指標として有効である一方、頭部の動きは「人の特性」を映し出す傾向が強いという結論に至った。

表 3 頭部位置の動きによるモードと楽曲ごとのユークリッド距離（昇順でソート）

mode	song	avg_6D_euclidean_distance
M15	S02	0.49
M09	S01	0.57
M08	S01	0.59
M15	S03	0.60
M10	S01	0.60
M01	S02	0.60
M09	S03	0.71
M10	S02	0.75
M03	S02	0.77
M01	S03	0.78
M07	S01	0.80
M10	S03	0.86
M08	S02	0.93
M03	S01	0.99
M09	S02	1.01
M03	S03	1.02
M07	S02	1.07
M01	S01	1.08
M15	S01	1.19
M14	S01	1.26
M08	S03	1.26
M14	S02	1.32
M14	S03	1.33
M07	S03	1.34

6.4 主観的評価との整合性の検証

定性的なインタビューから得られたユーザの主観的評価と、視線データの定量的特徴がどのように一致しているかを検討した。目的は、視線分析の信頼性と意味解釈の妥当性を高めることである。

参加者ごとにインタビュー記述から重視している体験の軸を抽出し、それが注視割合や遷移パターンと一致しているかを検証した。図 7 には、各参加者の代表的な発言内容と注視遷移ヒートマップを対応づけて示している。

P01 は「歌詞を裏から見る」「残された歌詞がウツとくる」といった発言をしており、OLO への注視が多く、OLO と CPO の間の遷移も頻繁であった。P02 は「今流れている歌詞に集中したい」「揺らめきがよかった」と語り、CCO と CPO の注視割合が高く、遷移も非常に少なかった。P03 は「歌詞をつまんだりして触った」「座り込んで見ていた」など操作性に言及しており、OLO と IVO への注視が多かった。P04 は構造的変化（落下・積み上げ・破壊）に関心を示し、OLO および TLO の注視と、広がった遷移構造が対応していた。

このように、主観的記述と視線データが整合している事例が多数確認され、視線行動に基づいた体験スタイル分析の妥当性が支持された。

6.5 PCA に基づくモード誘発性と個人差の関係

体験スタイルが参加者とモードのどちらにより影響され

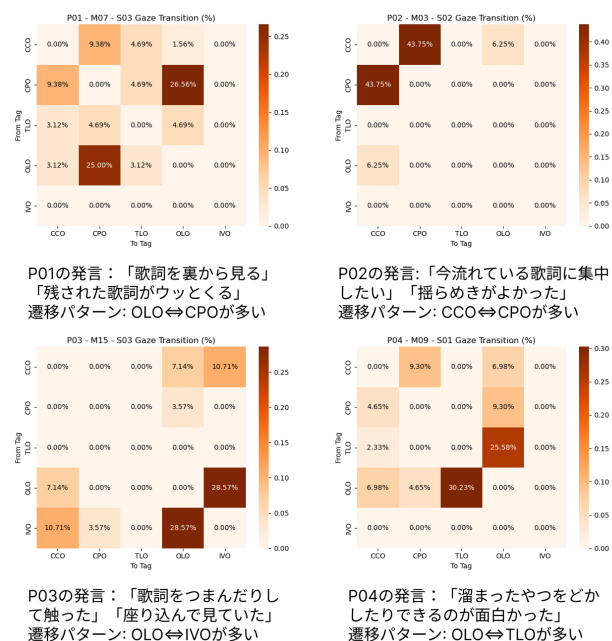


図 7 各参加者のインタビュー内容と注視遷移のヒートマップ

るかを明らかにするため、注視割合ベクトルを PCA により次元圧縮し、クラスタリングを行った。目的は、視線スタイルが参加者に固有のものであるのか、モードによって一貫して誘発されるのかを可視化することである。

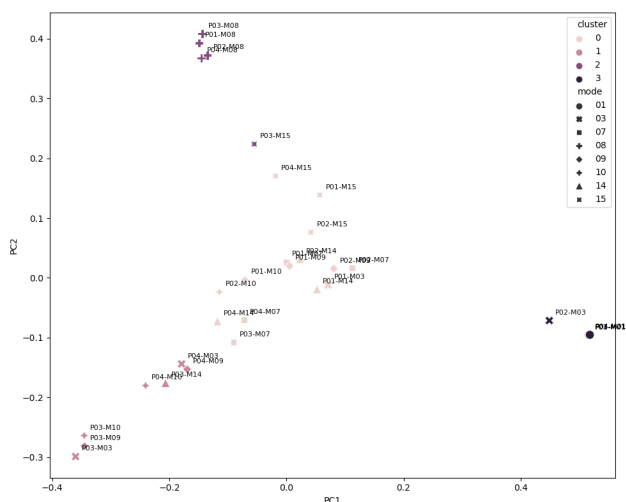


図 8 参加者ごとの各モードにおける注視遷移の PCA プロット

図 8 は、各試行の PCA プロットにクラスラベルを付与したものである。これを見ると、M01, M08, M07 ではすべての参加者が同一クラスに属しており、これらのモードでは視線スタイルがモードによって一貫して誘導されていることが示された。一方、M03 や M09 などでは参加者が異なるクラスに分散しており、モード内での視線行動が個人によって異なっていた。このように、モードごとに視線行動の誘発力に違いがあり、個人の重視する体験スタイルが表出しやすいモードと、モード設計によって体

験が固定化されるモードが存在することがわかった。

6.6 楽曲と体験スタイルの関係性の分析

最後に、楽曲の曲調が視線行動にどのように影響を与えているかを検討した。分析は 2 つの観点から行われた。

第一に、モードごとの「曲調に対する感受性」を測定するため、同一モードにおける 3 楽曲 (S01, S02, S03) の視線遷移行列のベクトルの分散合計 (total variance) を算出した。図 9 にその結果を示す。

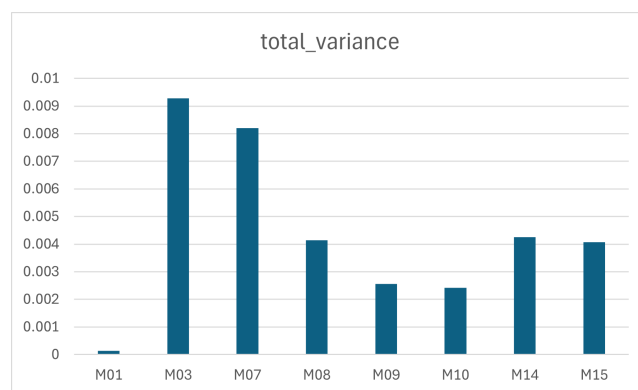


図 9 モードごとの total variance に関する棒グラフ

M03 は特に total variance が高く、曲調によって体験スタイルが切り替わりやすいモードであることが確認された。一方、M01 はどの曲調においても遷移構造が安定しており、曲調の違いによる視線行動の変化がほとんど見られなかった。

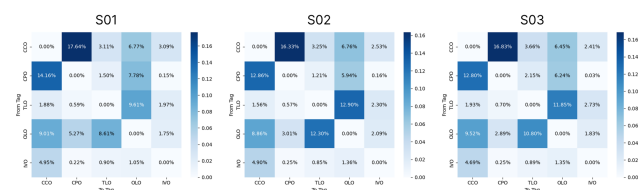


図 10 楽曲ごとの平均注視遷移率に関するヒートマップ

第二に、各曲調と体験スタイルの相性について、インタビューと視線スコアを用いて整理した (図 10)。S01 (ロック) は CCO/CPO への集中が高く、没入型の体験と親和性が見られた。S02 (ヒップホップ) は TLO/OLO が多く、操作的・探索的な体験が促されていた。S03 (バラード) は視線の分布が比較的バランスしており、複合型の体験が誘発されやすいことが示唆された。このように、モードだけでなく曲調も体験スタイルに影響を与える要因であることが明らかになった。モードのデザインと曲調の特性が互いに作用し合い、ユーザの体験を構成していることが、本研究からの重要な知見のひとつである。

7. 結論

本研究では、Mixed Reality 環境における視線と身体行動のデータを用いて、LyricBathe における歌詞体験の構造を明らかにした。分析の出発点として、「ある曲に適したモードがある」「モードによって誘発される体験が決まる」という2つの仮説を設定し、視線注視割合、注視遷移構造、頭部位置変化、インタビュー評価を総合的に分析した。注視割合の分析からは、モードによって視線行動が強く規定される場合と、個人差が顕著に現れる場合があることが明らかとなった。これは、モードの設計がユーザ体験における「誘発性」として機能していることを示している。また、頭部位置データの分析では、モードや曲調による影響は限定的であり、個人の身体的傾向がより強く反映されることが確認された。さらに、インタビュー結果と視線行動を照合したことで、主観的な体験評価と客観的な視線データが高い一致を示すことがわかった。これは、視線データに基づく体験スタイルの定量分析が、ユーザの主観的印象と対応していることを支持する。PCA とクラスターリングに基づく分析では、モードによって視線行動が収束する場合と、個人差がそのまま現れる場合が構造的に区別できた。また、楽曲の曲調によって体験スタイルが変化するモードも確認され、モードと曲調の組み合わせによって適切な体験が誘発されるという本研究の仮説を支持する結果が得られた。

これらの結果から、モードの設計は単に視覚的な演出にとどまらず、ユーザの体験スタイルそのものを形成する要素であることが明らかとなった。また、楽曲のテンポや構造といった属性に応じて最適なモードを推薦する適応的システム設計への展開可能性が示された。本研究は、視線行動の定量分析を用いて MR 環境における複雑なユーザ体験の構造を可視化する手がかりを提示するものであり、今後の体験設計における基盤的知見となると考えられる。

謝辞 本研究の一部は、JST、ACT-X、JPMJAX22A3 の支援を受けたものである。LyricBathe を体験し多くの貴重なフィードバックを下された方々に深く感謝する。

参考文献

- [1] Kato, J. and Goto, M.: Lyric App Framework: A Web-Based Framework for Developing Interactive Lyric-Driven Musical Applications, *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, Association for Computing Machinery, (online), DOI: 10.1145/3544548.3580931 (2023).
- [2] 小島颯英, 柏木敏朗, 中小路久美代: MR 環境での楽曲と同期した歌詞表示におけるユーザの注視箇所の分析, 情報処理学会研究報告 Vol.2024-HCI-209, 札幌, 北海道, 日本, 情報処理学会, pp. 1-7 (2024).
- [3] 小島颯英, 中小路久美代, 加藤 淳: MR 環境 LyricBathe におけるインタラクティブな歌詞表現のデザインスペース

- の創出, 信学技報 HCS2024-79,HIP2024-80 (2025-01), 茨木, 大阪, 日本, 電子情報通信学会, pp. 78-83 (2025).
- [4] Ford, S., Forlizzi, J. and Ishizaki, S.: Kinetic typography: issues in time-based presentation of text, *CHI '97 extended abstracts on Human factors in computing systems looking to the future - CHI '97*, Atlanta, Georgia, ACM Press, p. 269 (online), DOI: 10.1145/1120212.1120387 (1997).
- [5] Yokohama, H., Kasihwagi, T., Yamamoto, Y. and Nakakoji, K.: [How Different Text Animations Affect MR-Based Novel Poetry-Reading Experience] MR wo mochiita dokusho kankyou ni okeru moji no hyougen houshiki to shi no inshou no chigai ni kansuru kousatsu(in Japanese), *Human Interface Symposium 2022*, pp. 75-80 (2022).
- [6] Bandyopadhyay, P.: Immersive Space to Think: the Role of 3D Immersive Space in Sensemaking of Textual Data, PhD Thesis, Virginia Tech (2020).
- [7] Luo, W., Ellenberg, M. O., Satkowski, M. and Dachselt, R.: Documents in Your Hands: Exploring Interaction Techniques for Spatial Arrangement of Augmented Reality Documents, *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan, ACM, pp. 1-22 (online), DOI: 10.1145/3706598.3713518 (2025).
- [8] Amini, F., Riche, N. H., Lee, B., Monroy-Hernandez, A. and Irani, P.: Authoring Data-Driven Videos with DataClips, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 23, No. 1, pp. 501-510 (online), DOI: 10.1109/TVCG.2016.2598647 (2017).
- [9] David, Chung, S.-W. and Caires, C. S.: i-Cube, an (In)tangible Interface for Multi-user Generative Art Creation, *Entertainment Computing - ICEC 2023* (Ciancarini, P., Di Iorio, A., Hlavacs, H. and Poggi, F., eds.), Singapore, Springer Nature Singapore, pp. 187-192 (2023).
- [10] Miyagawa, S., Koyama, Y., Kato, J., Goto, M. and Morishima, S.: Placing Music in Space: A Study on Music Appreciation with Spatial Mapping, *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems*, Hong Kong China, ACM, pp. 39-43 (online), DOI: 10.1145/3197391.3205409 (2018).
- [11] Jeon, J. and Lee, C. H.: SoundMist: Novel Interface for Spatial Auditory Experience, *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, San Francisco CA USA, ACM, pp. 1-3 (online), DOI: 10.1145/3586182.3616622 (2023).
- [12] Jeon, J. and Lee, C. H.: Sprayable Sound: Exploring the Experiential and Design Potential of Physically Spraying Sound Interaction, *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan, ACM, pp. 1-17 (online), DOI: 10.1145/3706598.3713786 (2025).
- [13] Kato, J., Nakano, T. and Goto, M.: TextAlive: Integrated Design Environment for Kinetic Typography, *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, New York, NY, USA, Association for Computing Machinery, p. 3403-3412 (online), DOI: 10.1145/2702123.2702140 (2015).
- [14] Suno, Inc.: Suno. <https://suno.com/home> [2025-06-12].