

博士学位論文

舞台演劇と計算機の共創を支援する
インタフェースの提案

お茶の水女子大学大学院
人間文化創成科学研究科 理学専攻

西村 綾乃

2024年3月

要旨

オペラやミュージカルは、複数の創作領域を融合させた総合芸術である。創作においてはより良い成果を得るための試行錯誤が必須であり、舞台演劇においても稽古中に台詞や歌詞、演出の変更を行うことがある。

創作活動においては、小説やイラスト、音楽制作などに、計算機で動くソフトウェアが欠かせない存在となっている。計算機上で制作ができる利点として、何度もやり直しが出来る点や、同一作品の異なる形態を簡単に試作できる点が挙げられる。

舞台演劇における創作活動も、計算機と共創することでより良い作品を生み出す可能性がある。舞台で用いられる創作領域は多岐に渡るため、今回は、計算機が導入されていない訳詞と化粧に焦点を当て、共創の可能性を探求する。

創作補助としては、Large Language Model(LLM)の登場により、プロンプトを入力するだけで、質問に対する回答得たり、イラストを出力したりすることが容易になった。LLMなどを用いたチャットサービスや画像生成は汎用性があるためアイデアを出す段階での相談役として活躍する一方で、限定された状況での使用は困難な場合もある。そのため、条件や制約をあらかじめ考慮した上でユーザに生成内容を提示ができる、専用のインターフェースが必要であると考えた。

舞台芸術における創作環境に専用のインターフェースを提供することを考えた時、課題となる点が存在する。創作においては、より良いものを作成するために、幾度も修正を行い作品を磨き上げる試行錯誤の過程がある。そのため、計算機を用いた創作においても試行錯誤がしやすい、リアルタイムに編集状況が変化するインターフェースが必要である。そして、舞台に携わる人口はプロよりもアマチュアの方が多く、本提案の対象となる全てのユーザに専門的な知識があるわけではない。複雑なパラメータ設定や操作は困難であり、シンプルな操作で編集ができる必要がある。そこで、本論文ではシンプルな操作で創作活動が可能なユーザとの対話がリアルタイムに反映されるインターフェースを提案する。

訳詞においては、訳詞候補を選択し、ひらがなで編集すると、入力した訳詞がリアルタイムに譜面に反映されるインターフェースを提供する。その結果、ユーザが原言語や音楽に対する専門的な知識を持ち合わせていない場合でも、原曲の雰囲気を保ちつつ、歌詞の意味を反映できることを達成した。

化粧においては、アイシャドウ、アイライン、モチーフの特徴を反映したパーツを候補から選択、編

集することで、誰もが簡単に役に基づいた舞台用の化粧デザインが生成できるようなインタフェースを提供する。さらに、転写シールを使うことで個人の技量による仕上がりの差をなくしたメイクができるようになる。その結果、従来の化粧方法よりも舞台メイクに近いデザインが出来ることが確認された。

舞台演劇の分野に計算機を導入することで、試行錯誤がしやすくなり、アイディアの発散や新しい手法や表現の試作につながる。ひいては、演劇界が発展するとともに、舞台はより大衆化し、人々にとって身近な芸術となる。

abstract

Opera and musicals are comprehensive art forms that fuse multiple creative disciplines. Trial and error is an essential part of the creative process to achieve better results, and in stage plays, dialogue, lyrics, and direction are sometimes changed during practice.

Computer software is indispensable in creative activities; novels, illustrations and music compose. The advantages of working on a computer include the function to redo a work many times and the function to easily prototype different forms of the same work.

The creative activities of stage plays also have the potential to produce better works by co-creating with computers. In this thesis, we explore the possibilities of co-creation where computers have not yet been introduced; focusing on lyric translations and eye makeup.

Large Language Models (LLMs) have made a creative activity easy to obtain answers to questions and output illustrations simply by entering prompts. While chat services and image generation using LLMs are versatile enough to serve as a consulting tool during the idea generation stage, they can be difficult to use in limited situations. Therefore, a dedicated interface that can present the generated content to the user with conditions and restrictions taken into account in advance was considered necessary.

When considering the provision of a dedicated interface to the creative environment in the musical, there are some issues that need to be addressed. In creation, there is a process of trial-and-error in which the work is refined and revised many times in order to create a better product. Therefore, an interface that facilitates trial-and-error and changes the editing status in real time is necessary for computer-based creation. The population involved in musical productions is more amateur than professional, and not all users of the proposed system have expert knowledge. Complex parameter settings and operations are difficult, and editing must be simple. This paper proposes an interface that allows users to create music with simple operations and that reflects the user's interaction in real time.

For lyric translations, the proposed interface allows the user to select a candidate lyric translation,

edit it with hiragana, and have the input lyric translation reflected in the score in real time. As a result, even if the user does not have expert knowledge of the original language or the music, the system can reflect the meaning of the lyrics while preserving the atmosphere of the original song.

For makeup, we provide an interface that allows user to easily generate makeup designs for the stage based on the role by selecting and editing eye shadow, eye line, and parts that reflect the characteristics of the motif from a list of candidates. In addition, the use of fake tattoo stickers makes it possible to apply makeup without differences in the finished product depending on the skill of the individual. As a result, it has been confirmed that the design is more similar to stage makeup than conventional makeup methods.

The introduction of computers into the field of theatrical performance facilitates trial and error and leads to the divergence of ideas and the prototyping of new techniques and expressions. This will lead to the development of the theatrical field, and the stage will become a more popular and accessible art form for the public.

目次

第 1 章 緒言	1
1.1 背景	1
1.2 先行研究の現状と課題	2
1.3 本論文の貢献	4
1.4 本論文の構成	4
第 2 章 関連研究	6
2.1 はじめに	6
2.2 訳詞の先行研究の概要	6
2.2.1 歌詞翻訳と条件付き歌詞生成	6
2.2.2 韻律に基づいた歌詞やメロディの生成	7
2.2.3 字幕と文圧縮	8
2.2.4 先行研究の制限	8
2.3 化粧の先行研究の概要	8
2.3.1 メイクアップ技術の向上支援	9
2.3.2 入力画像を用いたメイクアップシミュレーション	9
2.3.3 ユーザの顔に変化を与える化粧表現の提案	10
2.3.4 先行研究の制限	10
2.4 文章や画像に関する機械学習	11
2.4.1 訳詞や化粧デザイン生成への応用	12
2.4.2 機械学習を用いた創作環境の課題点	15
2.5 むすび	15
第 3 章 訳詞アルゴリズムの提案と評価	16
3.1 はじめに	16
3.2 訳詞の課題	16

3.3	訳詞の手法	18
3.3.1	手法1	19
3.3.2	手法2	20
3.3.3	手法3	23
3.3.4	手法4	24
3.4	4つの手法の検討結果に対する考察	28
3.5	訳詞アルゴリズムの提案	29
3.5.1	原言語の翻訳	29
3.5.2	翻訳文の歌詞化	29
3.5.3	生成した歌詞のメロディへの当てはめ	31
3.6	訳詞アルゴリズムのモーラ誤差率による評価と結果	32
3.7	むすび	34
第4章	訳詞インタフェースの予備実装	35
4.1	はじめに	35
4.2	システムの予備実装	35
4.2.1	訳詞候補	36
4.2.2	歌詞編集	37
4.2.3	音符編集	37
4.3	インタフェース構成のための予備実験	38
4.3.1	訳詞候補の結果	40
4.3.2	歌詞編集の結果	41
4.3.3	音符編集の結果	42
4.3.4	アンケート結果	43
4.4	考察	44
4.5	むすび	45
第5章	訳詞インタフェースの実装と評価	46
5.1	はじめに	46
5.2	システム概要	46
5.3	システム実装	47

5.3.1	関連語の検索	48
5.3.2	訳詞の編集	48
5.3.3	試聴	48
5.4	手作業と提案システムによる訳詞作業の比較評価	48
5.4.1	結果	50
5.4.2	考察と議論	55
5.5	手作業による訳詞とシステムを用いた訳詞の印象評価	56
5.5.1	結果	57
5.5.2	考察と議論	58
5.6	モーラ誤差率と文章類似度による訳詞の定量評価	59
5.6.1	結果	60
5.6.2	考察と議論	60
5.7	まとめ	61
第 6 章	iMake: 役のモチーフに基づいたアイメイクデザインを生成するシステムの試作と予備実験	62
6.1	はじめに	62
6.2	舞台演劇におけるアイメイクの課題	62
6.3	試作システムのインタフェース	65
6.3.1	アイシャドウ	66
6.3.2	アイライン	68
6.3.3	アイペイント	68
6.4	試作システムによる予備実験と結果	68
6.4.1	考察	70
6.5	まとめ	70
第 7 章	iMake: 役のモチーフに基づいたアイメイクデザインを生成するシステムの実装と評価	72
7.1	はじめに	72
7.2	システム概要	72
7.2.1	アイシャドウ	73
7.2.2	アイライン	76
7.2.3	アイペイント	76

7.3	生成したデザイン3要素の印象評価	81
7.3.1	結果	82
7.3.2	考察と議論	84
7.4	従来の化粧方法とシステムを用いた化粧方法による化粧過程の比較実験	86
7.4.1	結果	88
7.4.2	考察と議論	94
7.5	従来の化粧とシステムを用いた化粧の印象評価	95
7.5.1	結果	96
7.5.2	考察と議論	97
7.6	まとめ	97
第8章	結言	99
A	：業績	102
B	：参考文献	104

目次

2.1	深山ら [Fukayama09] の提案した Orpheus のインタフェース.	7
2.2	阿部ら [Abe12] の提案した pâtisier のインタフェース.	7
2.3	[Tong07] らの化粧スタイル転送の例.	9
2.4	Kao ら [Kao16] のサーモクロミック顔料を用いたアイシャドウ.	10
2.5	Lulin ら [Lulin11] の LED を用いた表現.	10
2.6	[Qu20] らのダイナミックフェイシャルプロジェクションマッピング.	11
2.7	ChatGPT に原言語の歌詞とモーラ数を指定したプロンプトを入力した例.	12
2.8	モーラ数を減らすにはどうすべきかをプロンプトに入力した例.	13
2.9	同じプロンプトの入力に異なる回答が返ってくる例.	13
2.10	文化的文脈を反映しきれていない例.	14
2.11	リトルマーメイドのフランダールをモチーフにメイクデザインを指示した例.	14
2.12	closed eyes という単語を追加してデザインを出力した例.	14
3.1	メロディの高低と歌詞のイントネーション.	17
3.2	各手法のフロー図.	19
3.3	歌詞候補群選出の手順.	24
3.4	歌詞化の処理手順.	30
3.5	生成した歌詞のメロディへの当てはめの処理手順.	31
3.6	London Bridge の訳詞結果.	32
3.7	Under the Spreading Chestnut Tree の訳詞結果.	32
4.1	インタフェース選定用のシステム.	35
4.2	A0 と A1 のインタフェース.	36
4.3	B0 と B1 のインタフェース.	37
4.4	C0 と C1 のインタフェース.	38
4.5	各モジュールの平均作業時間と標準誤差.	39

4.6	A0 と A1 の比較結果.	40
4.7	B0 と B1 の比較結果.	41
4.8	C0 と C1 の比較結果.	42
4.9	システムに関するアンケート結果.	43
5.1	システム概要図.	46
5.2	システムインタフェース.	47
5.3	各機能の操作回数.	50
5.4	操作可能部分.	51
5.5	比較実験のアンケート結果.	52
5.6	訳詞の第三者による評価の結果.	57
5.7	質問 1 の平均値とモーラ数.	57
5.8	質問 2 の平均値とモーラ数.	58
5.9	ChatGPT に与えた入力と出力.	59
6.1	iMake を用いて転写シールで化粧をした例.	63
6.2	モチーフの特徴を取り入れたアイメイクデザインの 3 つの要素.	64
6.3	試作システムのインタフェース	65
6.4	画像の減色	66
6.5	色の抽出	66
6.6	グループに振り分けられた代表色	67
6.7	完成したグラデーション	67
6.8	アイペイントの生成	68
6.9	被験者によって生成されたデザイン	69
6.10	評価実験の結果 (質問 1-5)	69
7.1	システムインタフェース	72
7.2	アイシャドウのタブの内容	73
7.3	パレットの色数とグラデーション形状	74
7.4	モチーフ画像から抽出した 6 色	74
7.5	アイラインのタブの内容	76
7.6	アイペイントのタブの内容	77

7.7	アイペイント作成インタフェース	78
7.8	スタイル変換を利用したアイペイントへの模様付け	80
7.9	配色に対する回答者ごとの評価平均値	83
7.10	アイラインの色数評価結果 (グラフ)	84
7.11	アイペイントの形状と模様に対する回答者ごとの評価平均値	85
7.12	実験に用いた 2 種類のモチーフ画像 (出典: いらすとや)	86
7.13	(左) 従来のメイク方法による化粧, (右)iMake による化粧をした結果	88
7.14	iMake および従来のメイクの平均作業時間 (秒)	89
7.15	課題 1 に関する設問の結果	89
7.16	課題 3 に関する設問の結果	91
7.17	(左) 転写シールに関する設問の結果, (右) 目の形に関する設問の結果	93
7.18	質問 1 と質問 2 の平均値	96

表 目 次

3.1	単語の省略における類似度	22
3.2	単語の倒置における含有率	23
3.3	手法 3 による翻訳結果	23
3.4	Translation Result by Method Three.	23
3.5	実験条件	26
3.6	実験結果	27
3.7	訳詞アルゴリズムを適用した曲の RMSE.	32
4.1	各機能のモジュール内容	35
4.2	被験者の分類	38
5.1	各訳詞の RMSE および文章類似度.	60
5.2	定性評価で用いた評価対象の RMSE および文章類似度.	60
7.1	アイシャドウ配色評価結果	82
7.2	アイラインの色数評価結果 (数値)	83
7.3	アイペイント評価結果	84
7.4	システム評価実験の作業後アンケート内容	86
7.5	作業時間の結果	88
7.6	課題 1 に関する自由記述の回答	90
7.7	課題 2 に関する自由記述の回答	91
7.8	課題 3 に関する自由記述の回答	92
7.9	実験の感想	93

第1章 緒言

1.1 背景

舞台演劇には、オペラ、バレエ、ストレートの演劇、能、狂言、歌舞伎、ミュージカルなど、様々なジャンルが存在する。特にオペラやミュージカルは、音楽、舞踏、文学、演劇、映像など、複数の領域を融合させた総合芸術である。俳優や歌手、ダンサーなどのパフォーマンスも重要であるが、台本や演出、衣装、化粧やヘアメイクも舞台を構築する上で欠かせない要素となる。創作においてはより良い成果を得るための試行錯誤が必須であり、舞台演劇においても伝統的な作品で無い限り、稽古中に台詞や歌詞、演出の変更を行うことがある。

創作活動においては、小説やイラスト、音楽制作などに、計算機で動くソフトウェアが欠かせない存在となっている。舞台の場でも、3DCGを用いたプロジェクションマッピングなどが演出の一部として活用されている。計算機上で制作ができる利点として、何度もやり直しが出来る点や、同一作品の異なる形態を簡単に試作できる点が挙げられる。このように創作活動における計算機の利用は、成果物の質の向上に繋がる。

舞台演劇における創作活動も、計算機と共創することでより良い作品を生み出す可能性がある。舞台で用いられる創作領域は多岐に渡るため、今回は、計算機が導入されていない訳詞と化粧に焦点を当て、共創の可能性を探求する。訳詞の創作要素は、外国語の歌曲を日本語に訳詞するとき、言葉を選択したり、メロディを編集したりする部分にある。また、化粧では自分の役になりきるために、どのような色や形状のデザインにするかという点に創作要素がある。化粧については、計算機上のシミュレーションだけでは不十分であると考え、本論文では転写シールを用いて実際に化粧をする部分までを創作要素とし、人間の顔の魅力と最も相関があるとされている目に注目し [Saegusa15]、アイメイクをデザインの対象とする。

創作補助としては、Large Language Model(LLM)の登場により、プロンプトを入力するだけで、質問に対する回答得たり、イラストを出力したりすることが容易になった。一方で、機械学習を用いたサービスでは同じ内容をプロンプトに入力しても、出力が毎回異なる。創作において、ユーザの方針が明確な場合、毎回異なる結果が出力されることは、作業の妨げになり得る。訳詞や化粧には一定のルールが存

在する。例えば、訳詞では多くの場合、日本語に訳した際にモーラ数¹が増え、語彙の削減が必要になってくる。そのため、モーラ数を減らすために、類義語への置き換え、主語や助詞の削除、フレーズ間の意味の入れ替えなどが行われる。化粧では、アイシャドウを例にすると、目の縁から瞼の外側へ向かって薄い色へとグラデーションするように色を選択するなどのルールがある。以上を踏まえて、本論文ではルールベースでの実装を前提とする。

また、LLM などを用いたチャットサービスや画像生成は汎用性があるためアイデアを出す段階での相談役として活躍する一方で、メロディのような条件付きの翻訳や化粧単体のデザインなど限定された状況での使用は困難な場合もある。回答結果は学習内容に依存するため、チャットサービスでは音楽的なフレーズやリズム、発音を考慮できなかつたり、イラスト生成では転写シールに印刷する前提の目を瞑った状態の化粧画像の生成結果は崩れたような絵になったり、指定したモチーフの画像を学習していない場合はモチーフの特徴を反映することは不可能である。また、プロンプトエンジニアリングという単語が誕生したように、ユーザが求める回答を得るには、プロンプトにどのような入力をするべきかを学習する必要がある。舞台における訳詞や化粧のように制限がある創作物に対してテキストのみの入力、全ての条件を指定することは困難である。そのため、条件や制約をあらかじめ考慮した上でユーザに生成内容を提示ができるような、専用のインターフェースが必要であると考えた。

舞台芸術における創作環境に専用のインターフェースを提供することを考慮した場合、課題となる点が存在する。創作においては、より良いものを作成するために、幾度も修正を行い作品を磨き上げる試行錯誤の過程がある。そのため、計算機を用いた創作においても試行錯誤がしやすい、リアルタイムに編集状況が変化するインターフェースが必要である。そして、舞台に携わる人口はプロよりもアマチュアの方が多く、本提案の対象となる全てのユーザに専門的な知識があるわけではない。複雑なパラメータ設定や操作は困難であり、シンプルな操作で編集ができる必要がある。そこで、本論文ではシンプルな操作で創作活動が可能なユーザとの対話がリアルタイムに反映されるインターフェースの提供を目指す。

1.2 先行研究の現状と課題

舞台演劇の創作環境用のインターフェースを提案するにあたり、訳詞と化粧のそれぞれの先行研究の現状と課題を述べる。

計算機を用いた音楽に関する先行研究では、言語の韻律に基づいた歌詞やメロディの生成に関連する研究が数多く存在している。また、LLM による自然言語処理が主流になってきたことで、メロディやモーラ数を条件とした条件付き歌詞生成や歌詞翻訳の研究も増えてきた。歌詞やメロディの生成に関連する

¹一定の時間的長さをもった音の分節単位

研究ではユーザがパラメータを設定できるようなインターフェースも提供されている。一方で、歌詞翻訳の先行研究では、言語モデルを用いて出力した結果はテキストのみであり、メロディが付随していない。ユーザは自身で生成文とメロディを比較して、楽譜を編集する必要がある。また、歌詞やメロディ生成で提供されているインターフェースのパラメータ設定は複雑である。複数のパラメータに対して知識がないと正確な設定ができず、さらに、設定したパラメータはすぐに反映されず、結果を得るまでに待ち時間が発生しリアルタイムな編集はできない。

LLM を用いて訳詞を行った場合、プロンプトに条件を入力できるが、条件文を守った訳詞を出力できるとは限らない。また、同一の入力文に対して、毎回異なる出力が成されるのは、創作においてはユーザの負担の要因となる。加えて、文化的文脈の十分な反映が困難な場面も存在し、期待通りの翻訳結果を得ることが難しい場合もある。

モチーフに基づく化粧のデザインに関する先行研究においては、好みの化粧や顔写真の入力画像に近づくための化粧シミュレーションを出力するシステムと、他者の顔から化粧のみを転送しシミュレーションする技術が類似している。しかしながら、これらの研究では化粧の結果は画面上のシミュレーションに留まり、実際に化粧をする場合は出力された画像を見ながら、ユーザが自身の手で化粧をする必要がある。また、デザイン生成が目的である以上、要素ごとにデザインを編集できるインターフェースが必要である。入力画像を用いた先行研究では、出力がアイシャドウとアイラインが一体になっている場合がある。さらに、舞台上のメイクにおいては、役の特徴となるような形状や模様（魚であれば鱗、蜘蛛であれば蜘蛛の巣など）をアイメイクのデザイン要素として反映することがある。そのため、モチーフに合わせてこのような特徴となるパーツを編集できる必要がある。ペイントソフトのようなデザイン編集ソフトは自由度は高いが、モチーフの特徴の反映をユーザが手作業で行う必要があり、操作が煩雑な上に、デザインの技量が必要になるため、シンプルな操作とは言えない。

臉上に直接貼り付けるようなデバイスや、プロジェクタの追従によって顔を変化させる技術も開発されているが、舞台上での実用性には配慮されていない。デバイスは役者の顔の側面にコードを這わせる必要があり、これは見た目の美しさや役者の表現に悪影響を及ぼす可能性がある。さらに、舞台上でのプロジェクタによる追従には限界があり、実際の演技において制約を与える可能性がある。

画像生成 AI を用いたアイメイクのデザイン出力では、転写シールに印刷するための目を閉じた状態の化粧画像が生成しづらく、生成されたとしても、画像からメイク部分のみをトリミングして編集する必要がある。さらに、出力結果が毎回変化するため、好ましいとされるデザインに対する微調整（例：アイシャドウの濃さなど）が難しいと言える。

1.3 本論文の貢献

本論文では舞台演劇の創作環境として、シンプルな操作で創作活動が可能なユーザとの対話がリアルタイムに反映されるインタフェースを提案する。複雑なパラメータ設定を必要とせず、クリックやドラッグなどの簡単な操作を行うだけで編集を可能とし、創作に必要な基本部分はシステムが提供する。そして、インタフェースを用いることによって、ユーザが一定の質が担保された成果物を得られるようになることを目標とする。ルールベースの実装を採用することで、LLMのようなブラックボックスなシステムにおけるプロンプトエンジニアリングの必要性が排除され、ユーザは創作そのものに注力することが可能となる。

訳詞の創作環境に関するシステムでは、まず、インタフェースに提示するための訳詞の候補を出すためのアルゴリズムを提案する。その上で、訳詞の候補を選択し、歌詞をひらがなで編集するだけで、シンプルな操作で訳詞作業を実施できるようにインタフェースを設計する。また、入力した訳詞に合わせて譜割りを自動で行うことで、リアルタイムに譜面に編集状況が反映される。ユーザが原言語や音楽に対する専門的な知識がなくても、原曲のメロディを保ちつつ、日本語に訳した時に原言語の意味を反映できるようなシステムを目指す。

化粧の創作環境に関するシステムでは、役のモチーフに基づいたアイメイクデザイン生成を可能にするインタフェースを提供する。デザイン候補をクリックし、キャンバス上で移動や回転などのシンプルな操作を行うだけで、デザインを生成できるよう設計する。キャンバス上には編集内容がリアルタイムに反映される。さらに、生成されたデザインを転写シールに印刷し、実際のアイメイクとして利用できるようにすることを検討する。ユーザが選択したモチーフから、自動的にアイメイクのデザインに特徴を反映し、転写シールを使うことで個人の技量による仕上がりの差をなくした舞台メイクができるようになることを目指す。

舞台演劇において手作業で行われている創作領域に専用のインタフェースを提供することで、専門的な知識がなくとも一定の質を保った成果物を得られることができるようになる。そして、創作における試行錯誤がしやすくなり、アイデアの発散や新しい手法や表現を試しやすくなるだろう。

1.4 本論文の構成

以下に本論文の構成を述べる。

第2章では、舞台演劇の創作環境として計算機を用いたインタフェースを提案するにあたり、関連研究を紹介する。訳詞、化粧、機械学習の3種類に大別して先行事例とその制限を述べる。

第3章では、訳詞をする際の手法を4つ提案し、その実行結果と考察について述べ、訳詞のアルゴリズムについて説明し、その実行結果および評価実験に関する考察を述べる。続いて、第4章では、訳詞の作業を行うために適しているインタフェースを検討するために、各機能に対して2つの異なる実装を行い、任意に切り替えられるシステムについて説明し、予備実験と考察について述べる。第5章では、改良したインタフェースの実装について説明し、提案システムが訳詞における課題を克服できているかを3つの実験により評価し、その結果と考察について述べる。

第6章では役のモチーフに基づいたアイメイクデザインを生成する上で、提案インタフェースの有効性や転写シールによる化粧の実現可能性を検証するために試作を行った内容と予備実験について述べ、第7章では改良したインタフェースの実装について説明し、提案システムが生成したデザインに対して印象を評価する。さらに、提案システムと従来の化粧法の比較実験とその結果および考察について述べ、比較実験で得た化粧写真が舞台に近いメイクが出来ているか評価することで、化粧における課題を克服できているかを考察する。

最後に、第8章で総括することで、本論文をまとめる。

第2章 関連研究

2.1 はじめに

本章では、舞台演劇の創作環境としてのインタフェースを提案するにあたり、先行研究を紹介する。2.2節では訳詞に関する先行研究について述べ、2.3節では化粧に関する先行研究について述べる。最後に2.4節にて、機械学習技術の先行研究の紹介と本論文で提案するシステムへの応用について議論を行う。

2.2 訳詞の先行研究の概要

計算機を用いた歌詞やメロディの生成に関する研究は数多く成されている。2.2.1節では、歌詞翻訳やモーラ数などの条件付きの歌詞生成に関する研究について述べる。2.2.2節では、言語の韻律に基づいた歌詞やメロディの生成に関連する研究について述べる。

また、映画やテレビ字幕の翻訳には、歌詞の翻訳と似たような問題がある。字幕は表示される時間が短いため、観客がそれを知覚し理解できる時間は限られている。この問題は、歌詞翻訳におけるモーラ数の制限と似ていると考えている。さらに、翻訳した文章のモーラ数を減らす課題は、自然言語処理における文圧縮の問題と関連がある。2.2.3節では、字幕と文圧縮に関連した研究について紹介する。最後に、2.2.4節にて、先行研究の制限についての議論を述べる。

2.2.1 歌詞翻訳と条件付き歌詞生成

吉村は、既存のニューラル機械翻訳と歌詞用に学習した事後編集器を組み合わせた歌詞翻訳のシステムを提案している [Yoshimura20]。このシステムは、原文の歌詞をニューラル機械翻訳にかけて得られた訳文に対して、歌詞翻訳用に学習させたニューラル機械翻訳で言い換えを行い、最後にモーラ数や文末表現を調整することで訳詞を出力している。櫻井らは、ベースとなるバイリンガルLLMに対して Reinforcement Learning from Human Feedback (RLHF) を用いたモーラ数制御を提案している [Sakurai23]。

また、歌詞翻訳とは異なるが、メロディ条件付きの歌詞生成の研究も存在する [watanabe18]。歌詞の意味に対する制約は無いが、メロディ条件がつくことによりモーラ数を制御しながら歌詞を生成するという点において、訳詞と近い課題設定となっている。

2.2.2 韻律に基づいた歌詞やメロディの生成

はじめに、韻律に基づき自動で歌詞やメロディを生成する研究について述べる。メロディと歌詞をマッチさせるための韻律の要素には、音節、強拍弱拍、モーラ数、母音の位置などが挙げられる。Tra-la-lyrics 2.0[Gonçalo15]は、Tra-la-lyrics[Gonçalo07]とPoeTryMe[Gonçalo14]を統合した、入力されたメロディに基づき、歌詞の音節と強拍の位置をメロディとマッチさせ、さらに一貫した意味を持つ歌詞を自動生成するシステムである。強拍弱拍の特定には、Lerdahl らの提案したドットシステムが使用されている[Lerdahl85]。

伊藤らは入力文のモーラ数を算出し、その結果から入力文を歌詞として割当可能な楽曲リストを作成し、ユーザに提示する、暗記のための替え歌を自動生成するシステムを提案している[Ito15]。このシステムは、提示されたリストからユーザが曲を選択すると、モーラ数とメロディの音符数が完全一致しない部分について音符に変更を加えることで、入力文を歌詞とした歌曲を生成する。歌詞作成支援システム「LyriSys」は、節やサビなど歌詞の構造を総合的に考慮した作詞支援システムである[Watanabe17]。モーラ数を含む楽曲の構造とストーリー要素、形態素パターンを指定すると、ストーリーを考慮した歌詞や、1番と2番で類似した形態素パターンを持つ歌詞を生成する。



図 2.1: 深山ら [Fukayama09] の提案した Orpheus のインタフェース。



図 2.2: 阿部ら [Abe12] の提案した pâtisier のインタフェース。

生成できる歌詞やメロディを調整できるインタフェースを提供しているシステムには、Orpheus がある[Fukayama09](図 2.1)。これは、入力された日本語の歌詞のイントネーションを考慮したメロディを生成する自動作曲システムである。日本語においては「雨」と「飴」など同じ言葉でもイントネーションで意味が変わってしまう場合がある。オルフェウスではこのような歌詞のイントネーションを解析し、動的計画法を用いて歌詞に相応しいメロディを探索する。また、阿部らによる「pâtisier」は、音節数や韻、単語のアクセントに着目した作詞支援システムである[Abe12](図 2.2)。Ngram 言語モデルをもとに、ユーザの指定した条件を満たす歌詞候補文を生成する。指定できる条件は、文頭の 2 単語、モーラ数、母

音, アクセントなどである.

2.2.3 字幕と文圧縮

文書翻訳においては, 一般的な文章だけでなく, 漫画や字幕など様々な場面を想定した機械翻訳システムが考案されている.

字幕に関する研究では, Popowich らが北米のテレビ放送における英語からスペイン語への字幕翻訳システムを提案している [Popowich00]. 話し言葉が多くなること, また, 画面に表示される時間が数秒のため翻訳を理解する時間が限られていることの2点を字幕の特徴として挙げている. また, Volk は機械翻訳の成功の指標として, ユーザ数, 満足度, 翻訳システムの使用期間の3つを挙げ, 映画字幕においてはどの指標が当てはまるかについて議論している [Volk08].

文圧縮の分野では, 井手上らは, 適当に要約しながら翻訳する場面を想定し, 部分的表現を中心として翻訳を行う部分的機械翻訳 [Yamamoto13] を提案している. この手法は, 原言語文と原言語文の一部の表現 (部分的表現) を入力とし, 部分的表現を中心とした簡潔な翻訳を出力するものである. また, 新幹線車内の電光掲示板で流れるような簡潔な表現のために, 体言止めや助詞止めを適用する研究も存在する [Yamamoto05].

2.2.4 先行研究の制限

2.2.1節では, 歌詞翻訳や条件付きの歌詞生成について紹介した. しかしながら, 現状での訳詞研究は言語モデルによる出力に留まり, メロディが付随しない. そのため, ユーザは自身で生成文とメロディを比較して, 楽譜を編集する必要がある. 歌詞またはメロディの一方を生成する研究は, 2.2.2節で述べた内容を含め数多く存在するが, 本論文の目的は, 歌詞とメロディの両方が事前に提供されているという前提の下で, 形式と内容を維持しながら歌詞とメロディを生成することである. 提供されているインタフェースには多くのパラメータを設定する項目があり, 中には専門的な知識が無ければ正確な設定が難しい項目も存在する. また, 設定したパラメータはすぐに反映されず, 変更した結果を得るまでに待ち時間が発生する. 2.2.3節で述べた, 映画字幕の翻訳では, 決められた時間内に理解できる文章を提供する点, 文章圧縮では, モーラ数が減るという点において, 訳詞作業との関連がある.

2.3 化粧の先行研究の概要

化粧に関連する研究は以下の3つに分けて説明を行う.

- メイクアップ技術の向上支援

- 入力画像を用いたメイクアップシミュレーション
- ユーザの顔に変化を与える化粧表現の提案

2.3.1 メイクアップ技術の向上支援

Lui らの“Beauty e-Experts”[Liu13] や, Wang らが提案したシステム [Wang06] はユーザの写真から顔のパーツを認識・検出し, ユーザの顔の上にメイクを施した様子を画面上に提示する. 商品として存在する化粧品の情報を用いて, ユーザのメイクアップ支援を行う研究も存在する. Rahman らの提案したスマートミラーシステムは化粧品につけた RFID を読むことで, その製品を使って化粧した結果を, カメラで撮影しているユーザの顔に対してシミュレーションする [Rahman10]. また, 中川らの提案した“Smart Makeup System”は, ユーザが化粧をしたときの写真と使用した化粧品情報の記録を Web 上で他者と共有することで, 化粧のバリエーション増加を支援するシステムである [Nakagawa11].

非日常的な化粧支援に関する研究では, 中国の代表的な伝統音楽劇である京劇の化粧デザインを目, 鼻, 口などの部位に分けて組み合わせ, インタラクティブかつリアルタイムに新しい化粧のパターンを 3D レンダリングするシステムがある [Cai10].

2.3.2 入力画像を用いたメイクアップシミュレーション



図 2.3: [Tong07] らの化粧スタイル転送の例.

神武らは好みの顔画像とユーザ自身の顔画像を入力することで, 好みの顔画像にユーザが近づくためのメイクアップシミュレーション画像とメイクアップ商品の候補を提示するシステムを提案している [Kotake17]. メイクアップ画像からメイクのスタイルを抽出し, 非メイクアップ画像にスタイルを移し替える研究は多数存在し, ニューラルネットワークを用いた手法 [Organisciak21][Zhu19] だけでなく, ピクセル単位で画像の処理を行いマッピングすることで化粧スタイルを転送する手法もある [Tong07] (図 2.3).

2.3.3 ユーザの顔に変化を与える化粧表現の提案

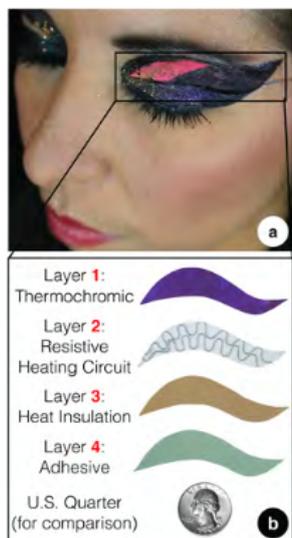


図 2.4: Kao ら [Kao16] のサーモクロミック顔料を用いたアイシャドウ.

従来のメイク方法にとらわれない、新しい化粧の表現もいくつか提案されている。“ChromoSkin”は電気制御や周囲の温度条件によって変化するサーモクロミック顔料で作られたアイシャドウタトゥーを用いて、インタラクティブに色を変化させる化粧を提案している [Kao16] (図 2.4)。“Digital Eyeshadow”は、目を閉じたときに顔の化粧の全貌が見えることに着目し、LED を用いた表現を提案している [Lulin11]。顔の上に複数の LED を装着し、目を閉じたり瞬きしたりすることで、LED が発光して顔の上にさまざまなパターンが浮かび上がる (図 2.4)。

ユーザの顔に特殊な機器を取り付けることなく変化を与える研究では、高速カラープロジェクタを用いて直接映像を投影するダイナミックフェイシャルプロジェクションマッピング [Qu20] があり、顔の動きや表情の変化にもリアルタイムで追従する (図 2.6)。この研究をメイクシミュレータとして発展させた“COLOR MACHINE” [Peng23] は、アイシャドウ、チーク、リップによるメイクのシミュレーション結果をユーザの顔に立体的に追従して投影する。

2.3.4 先行研究の制限

化粧の研究領域では、2.3.1節で述べたように、技術向上やデザインのバリエーションを増やす研究は多数存在するが、画面上でのシミュレーションに留まる。2.3.2節で紹介した入力画像に基づいたメイクアップシミュレーションシステムは、入力画像を用いる点、役のモチーフに基づくという点と類似しているが、出力画像のアイシャドウとアイラインが一体化している場合がある。デザインの生成が目的



図 2.5: Lulin ら [Lulin11] の LED を用いた表現.



図 2.6: [Qu20] らのダイナミックフェイシャルプロジェクションマッピング。

である以上、アイシャドウとアイラインは要素ごとに調整できるべきである。また、役に基づいたアイメイクのデザインを考える場合、モチーフの形状のような特徴をアイメイクのデザイン要素として反映する場合がある。そのため、モチーフごとに形状や模様などを編集できる必要がある。デザインを編集するためのインターフェースとしては、ペイントソフトが存在する。しかしながら、編集の自由度は高い一方で、モチーフの特徴の反映をユーザが手作業で実施する必要があるため、操作が煩雑な上にデザインをするためのスキルが必要になる。

2.3.3節で述べた研究では、実世界でユーザの顔に変化を与えることができるが、舞台上で使用することは考慮されていない。顔の側面にコードを貼る必要があるため、役者の見た目を損なう可能性がある。また、プロジェクタは舞台上での追従に限界があり、役者の演技を阻害する可能性もある。

2.4 文章や画像に関する機械学習

Transformer は、Vaswani らによって提案された Attention のみで構築されたニューラルネットワークアーキテクチャである [Vaswani17]。それまで、文章を処理するタスクでは Attention と同時に Recurrent Neural Network(RNN) を使用するのが主流であった。Transformer の登場により、Bidirectional Encoder Representations from Transformers(BERT)[Devlin18] や GPT-4[Achiam23] などの大規模言語モデルの構築が可能になった。特に、GPT-4 は翻訳や質問応答などの自然言語処理に関わるタスクだけでなく、入力画像からテキスト出力ができるマルチモーダルなモデルである。また、Transformer は文章に関連するモデルだけではなく、Latent Diffusion Models(LDM)[Rombach22] のような、テキストからの画像

生成や超解像度化などが可能な画像処理に関するモデルにも使用されている。

2.4.1 訳詞や化粧デザイン生成への応用

文章や画像を処理できる機械学習モデルは、訳詞やアイメイクデザインを考える際に有用である。例えば、訳詞であれば、モーラ数を減らすために類義語の候補を出してもらったり、翻訳の時点でモーラ数を指定して母国語に訳すよう指示することも可能である。アイメイクのデザインであれば、演じる役のモチーフとなるイメージ画像をテキストの入力から生成してもらうことや、役の特徴を抽出したアイメイクデザインを生成してもらうことも出来るだろう。

しかしながら、機械学習を用いる際は、既存のモデルを使用して、転移学習やファインチューニングを行う場合でも、訓練に必要な十分な量のデータを集める必要がある。今回の目的は舞台演劇における創作環境のインタフェースの提案であるが、訳詞と化粧においては、原言語と日本語の両方の譜面データや、転写シールの使用を前提としたモチーフに基づいたアイメイクデザインの画像の収集は非常に困難である。また、LLM を用いた機械学習モデルではプロンプトの入力次第で得られる回答が異なる。プロンプトエンジニアリングという言葉が誕生したように、ユーザが望む回答を得るためにはプロンプトを使いこなすためのスキルが必要になってくる。

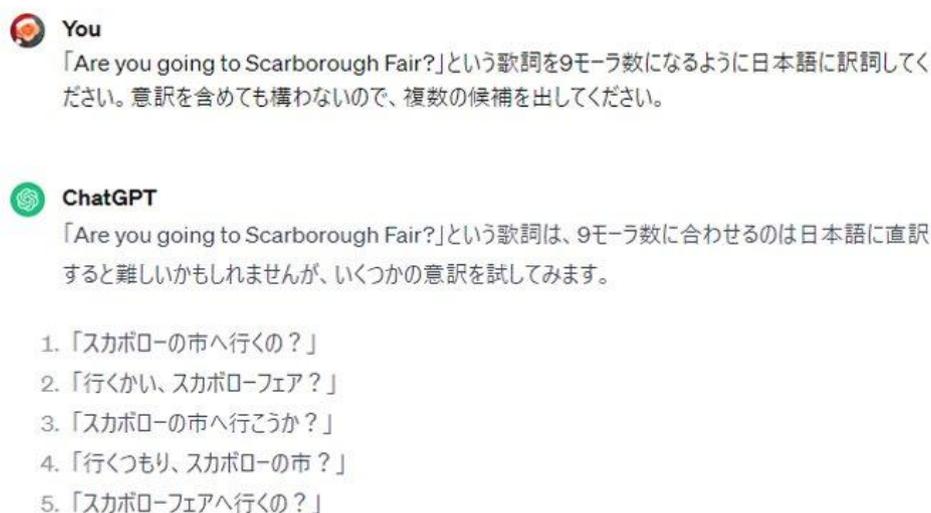


図 2.7: ChatGPT に原言語の歌詞とモーラ数を指定したプロンプトを入力した例。

ChatGPT¹や StableDiffusion²などの Web サービスの登場により、これらのツールを使用しつつ創作を行うことが一般的になってきたが、訳詞やモチーフを用いた化粧のデザインなど特定の領域への利用はファインチューニングなしでは難しいと考えられる。

¹<https://chat.openai.com/>

²<https://ja.stability.ai/stable-diffusion>

You
どうしたら9モーラ数に収まりますか？

ChatGPT
「Are you going to Scarborough Fair?」という英語の歌詞は、直訳すると日本語にすると9モーラに合わせるのが難しいです。しかしながら、意識を行うことでモーラ数を調整することができます。

以下はいくつかの意識例ですが、日本語において9モーラに完全に合致させるのは難しいため、ある程度の意識を含んでいます。

1. 「スカボローフェアへ行くのか？」 (8モーラ)
2. 「スカボローフェアに行く？」 (7モーラ)
3. 「スカボローフェア、行くの？」 (7モーラ)

9モーラに厳密に合わせることは難しいですが、9モーラ以内でより近い意味を保ちつつ意識することができます。

図 2.8: モーラ数を減らすにはどうすべきかをプロンプトに入力した例.

You
「Are you going to Scarborough Fair?」という歌詞を9モーラ数になるように日本語に訳詞してください。意識を含めても構わないので、複数の候補を出してください。

ChatGPT
「Are you going to Scarborough Fair?」を9モーラ数になるように訳詞するために、意識を含めた複数の候補を以下に示します：

1. 「スカボローフェア行くのかい」
2. 「スカボローフェア行くかい」
3. 「スカボローフェアへ行くのか」
4. 「スカボローフェアへ行くかい」
5. 「行くのかい、スカボローフェア」

これらは意識を含めても9モーラ数になるような訳詞の一例です。

You
「Are you going to Scarborough Fair?」という歌詞を9モーラ数になるように日本語に訳詞してください。意識を含めても構わないので、複数の候補を出してください。

ChatGPT
「Are you going to Scarborough Fair?」という歌詞は、9モーラ数に合わせるのは日本語に直訳すると難しいかもしれませんが、いくつかの意識を試してみます。

1. 「スカボローの市へ行くの？」
2. 「行くかい、スカボローフェア？」
3. 「スカボローの市へ行くかい？」
4. 「行くつもり、スカボローの市？」
5. 「スカボローフェアへ行くの？」

図 2.9: 同じプロンプトの入力に異なる回答が返ってくる例.

図 2.7に ChatGPT 3.5 に原言語の歌詞とモーラ数を指定し、日本語に訳すように指示した例を示す。訳詞の候補が5つ表示されているが、すべて9モーラ以上となっており、条件を満たしていない。図 2.8で、9モーラに収まるにはどのようにすべきかプロンプトに入力したところ、9モーラに収まる訳詞の候補も存在するが、候補の右に表示されているモーラ数は異なる表記となっている。

図 2.9では、同様のプロンプトの入力に異なる回答が返ってきた例である。機械学習を用いたモデルは、このように同じ入力でも違う出力を返すランダム性があるため、創作物がある程度明確になっている段階においてはユーザーにストレスを与える原因にもなり得る。

図 2.10に原言語の歌詞を日本の文化的文脈を考慮して意識し、指定のモーラ数になるようプロンプトに入力した例を示す。ここで入力した歌詞は本来、魔除けの意味を含むものであるが、ChatGPT の意識

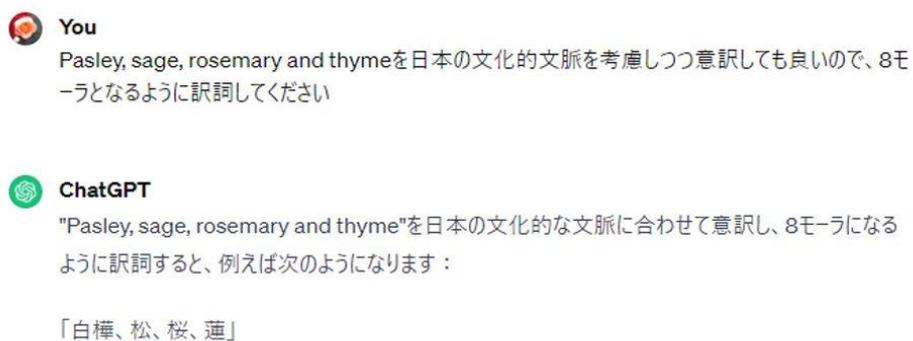


図 2.10: 文化的文脈を反映しきれていない例.



図 2.11: リトルマーメイドのフランダールをモチーフにメイクデザインを指示した例.



図 2.12: closed eyes という単語を追加してデザインを出力した例.

では訳しきれていない。また、モースラ数も指定した数とは異なる。

以上のようにモースラ数を指定した訳や文化的文脈を反映させることは難しく、訳詞用にチューニングする必要があると考えられる。

また、Hugging Face³にて提供されていた Stable Diffusion 2.1 を用いてモチーフを指定したアイメイクのデザイン生成を行った(図 2.11)。プロンプトの入力は“eyemakeup, musical makeup, Flounder motif, Little Mermaid, face sticker design”とし、リトルマーメイドのフランダールをモチーフにメイクのデザインを指示した。転写シールに印刷するべきデザインは目を閉じている状態のアイメイクデザインであるべきだが、プロンプトに“closed eyes”の単語を追加すると、化粧のデザインとは関係ない画像が出力される頻度が増える(図 2.12)。そのため、今回は“closed eyes”という単語を省略した。

³<https://huggingface.co/>

図 2.11の右端のデザインはモチーフの色や形状を表現できているように見えるが、目を開けた画像であるため、そのまま使用することは難しい。また、図 2.11の左端のデザインは目を閉じた画像として出力されているため、目の周囲の化粧を切り抜けば印刷が可能であるが、まつ毛の下のデザインが隠れているため手作業での補修が必要である点と、切り出した画像の縁の部分はグラデーションをかけないと肌に馴染まない点の2点が問題点として挙げられる。そのため、今回提案する化粧システムの趣旨とは異なる部分での課題が増える。

2.4.2 機械学習を用いた創作環境の課題点

機械学習を取り入れることは、システムにとって非常に有用であるが、モデルの性能がデータの量や質に左右されたり、プロンプトの入力が同一でも得られる結果が変わってしまうという欠点がある。訳詞やアイメイクにはルールが存在し、ルールベースでのアプローチは大量のデータやデータの品質に依存しないシステム構築が可能である。そのため、今回は専用のインタフェースを提供することを前提として、訳詞およびアイメイクのルールを模索し、定義した上でシステム構築を行う。

また、機械学習を用いた創作環境では汎用性はあるものの、限定された条件下での創作環境としての使用は困難であると考えられる。プロンプトに細かな条件を入力することは可能だが、条件を全て守った出力になるとは限らない。学習データが少なかったり、そもそも学習データとして存在していなかったものを出力することは困難である。さらに、入力はテキストに限られるため、全ての指定を言葉で行う必要がある。しかしながら、専用のインタフェースがあれば、あらかじめ、条件や制約を考慮した上でユーザに生成内容を提示することが可能である。

2.5 むすび

本章では、舞台演劇の創作環境としてのインタフェースを提案するに際し、訳詞(2.2節)と化粧(2.3節)に関連した先行研究とその制限について述べた。これにより、それぞれの領域に特化した専用インタフェースの提案において、先行研究の手法では課題を十分に克服できていないことが明らかになった。

さらに、2.4節では、機械学習を舞台演劇の創作環境に用いる場合の議論を述べた。機械学習を活用した創作環境は汎用性が高い一方で、特定の条件下では課題が多く残ることを指摘した。そのため、専用のインタフェースを提供することが、条件や制約を考慮した上でユーザに適切な創作補助を行うために有益であるとの結論に至った。

第3章 訳詞アルゴリズムの提案と評価

3.1 はじめに

本論文で提案する訳詞システムは、原言語 (英語) を日本語に翻訳し、訳詞の候補を自動で出力する。本章では、3.2節で訳詞の課題を挙げ、3.3節にて、訳詞アルゴリズムの実装に向けた訳詞の4つの手法について調査と考察を行った結果を述べる。3.5節では提案するアルゴリズムの手順を示し、3.6節ではアルゴリズムをモーラ数の増減にもとづいて評価する。

3.2 訳詞の課題

はじめに、舞台演劇の訳詞を補助するためのインタフェースを提供する上で、課題となる点を述べる。

ミュージカルのオリジナル作品は日本語でない場合が多いため、国内で上演するためには訳詞の作業が必要になる。歌詞の言語が曲のリズムに影響を与えるという研究も存在し [Patel03][Sadakata03]、言語の特徴や発音の違いが、曲のリズムやフレーズに合わせる際の調整に影響を与えることが示唆されている。訳詞では単に原言語の歌詞を日本語に翻訳しモーラ数を合わせるだけでなく、メロディのフレーズやリズム、さらには歌唱時の発音を考慮する必要がある。そのため、言語と音楽の両方の知識が必要になる。しかしながら、1.1節で述べたように、本論文で提案する創作環境の対象になる多くのユーザは言語と音楽の両方の専門的な知識は持ち合わせていないと考えられる。そのため、訳詞をした際に、原曲の雰囲気を保てず、歌詞の意味を反映できないという課題がある。

次に、外国語の歌曲を日本語に訳する場合に、どのような課題点があるかを述べる。

日本語への訳詞の歴史は近藤朔風らが切り拓いた [Matsuda07]。近藤は原言語の意味と韻律を可能な限り日本語に再現しようとする訳詞を重視した。

言語は一般に、ストレスタイム言語、シラブル言語、モーラ言語の3種類に分類される。日本語はモーラ言語に属している [Port87][Ramus02]。本研究では、1つのモーラを母音、子音に続く母音または拗音と定義する。拗音は、任意の「イ」母音を持つ発音の後に、小さなサイズの「ゃ」「ゅ」「ょ」が続くことを示す。原言語の歌詞を日本語に訳すと多くの場合、音数 (モーラ数) が増えるという問題点に直面する。メロディに使われている音符の数は決まっているため、訳詞においてはモーラ数を極端に増やすこ



図 3.1: メロディの高低と歌詞のイントネーション.

とは避けなければならない。結果として日本語への訳詞において、原言語の意味を省略する必要がある [Matsuda13].

歌唱に適した日本語の発音についても課題が指摘されている [Yamamoto73]. 音節の観点から、撥音、促音、長音、二重母音と呼ばれる発音を、モーラ数を1とする場合と、直前の発音と合わせて歌うことで、モーラ数に含めない場合がある。これは、歌唱におけるリズム感に影響を与える。

母音の観点からは、松田は本文中で、こうも述べている。“曲の終わりに歌いあげる音では、意味のレベルを犠牲にしても「ア」か「オ」の母音にすべき場合がある。”これは、歌唱に適した日本語の発音の課題とも言える。山本は日本語の母音について以下のように述べている。“「あ・お」は会話にも歌唱においてもひびきやすい母音であるが、「え」は歌唱においてとくにひびかせにくい母音である。そこでたえず発声訓練において美しい共鳴をとるよう配慮されている。「い」「う」は歯をくいしばったように口腔中が狭くなり、共鳴のない固い平たい声になりやすいのである。”以上のことから、訳詞を行う際に、選択した単語の母音に対して、言葉の意味を重視するか、歌唱的な面を重視するかといった訳詞の問題が生じることが分かる。

加えて、訳詞家2名に、訳詞の課題点についてどのような点があるかヒアリングを行った。得られた意見を下記に示す。

1. 原曲の言語が英語の場合、1音が1単語に対応すると仮定すると、歌詞を日本語に翻訳した後、モーラ数が増えるため、歌詞をメロディに合わせるのが難しくなる。
2. フレーズの最後の重要な単語が高音やロングトーンである場合、母音の観点からメロディに適した日本語を選択することが難しい。
3. 日本語は「私」のような主語が無い曖昧な文章でも完結するので、モーラ数を減らせる場合が多い。
4. フレーズ冒頭の冠詞 (“a” や “the”) は、アウフタクト¹であることが多く、後に重要な言葉が続く場合に、アウフタクト部分の歌詞を感嘆詞 (“ああ” や “おお”) で代用すると、歌詞の表現が稚拙に見

¹第1小節目の前に配置された弱拍の音符のこと

える.

5. 歌詞を日本語に翻訳する際、1つの音符に1つのモーラしか割り当てないのが普通であるため、歌唱時において、歌詞の単語が正しく聴き取られるかどうかは課題である

5について、具体的な例として、同じ発音であるが音の高低が異なる単語（例：「雨」と「飴」）が存在することが挙げられる（図3.1）。この場合、メロディの高低が言葉のイントネーションと異なる場合、正しい単語として聞き取られない場合がある。そのため、訳詞を行ったときには日本語の韻律を考慮してメロディの高低への当てはめたり、休符の位置を考慮する必要がある。

ここで、訳詞をする際に課題となる点をまとめる。

課題 1 日本語への翻訳によってモーラ数が増える

課題 2 原言語の意味を取捨選択する必要がある

課題 3 撥音などの発音を1音とするかどうか、リズムと音節を考慮して判断する必要がある

課題 4 音楽的な理由から、特定の音に対する母音を限定する必要がある

課題 5 日本語のアクセントとメロディの高低を合わせて単語が正しく聴き取れるようにする必要がある

これらの課題に対して条件を明確にし、ルールを定めることで解決を試みる。そうすることで、システムが提示する訳詞の品質を一定に保つことが可能になる。本論文では、課題1を解決することに焦点を当てつつ、ユーザが原言語や音楽に対する専門的な知識がなくても、原曲のメロディを保ちつつ、日本語に訳した時に原言語の意味を反映できるようなシステムの構築を目指す。

3.3 訳詞の手法

前節で列挙した課題を踏まえて、訳詞のための和訳手法について考察する。なお実験では、使用する歌曲の原語を英語に限定する。

課題1に示した通り、日本語への翻訳ではほぼ間違いなく、モーラ数を減らす必要が生じる。この解決手法として、自然言語処理における文圧縮の問題が参考になると考えた。翻訳に関係する文圧縮の手法として井手上らは、部分的表現を中心として翻訳を行う部分的機械翻訳 [Ideue13][Yamamoto13] を提案している。この手法は、原言語文と原言語文の一部の表現（部分的表現）を入力とし、部分的表現を中心とした簡潔な翻訳を出力するものである。また、日本語文の要約に関する分野では、要約したい文より必要な単語群を抽出し、それらを入力とした上で不足する語を補い、文生成を行うことで、要約された文章を作る研究 [Ikeda06] も発表されている。

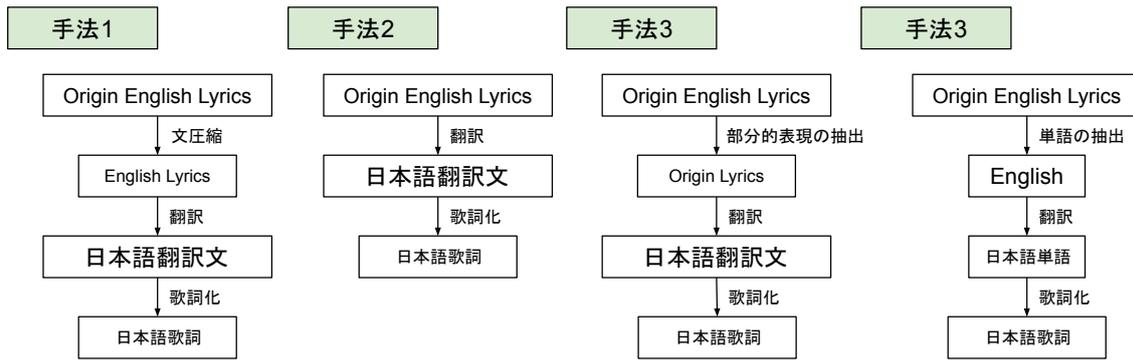


図 3.2: 各手法のフロー図.

これらを参考にして、歌曲における原語から日本語への訳詞の手段として、以下の4つの手法について検証・考察することにした。各手法のフローについて図 3.2に示す。

手法 1 メロディの1フレーズに該当する原文を圧縮して日本語に翻訳し、その文章を歌詞に最適な形にする

手法 2 原文に変更を加えず日本語に翻訳し、その文章を歌詞に最適な形にする

手法 3 原文から部分的な表現を抽出して日本語に翻訳し、生成された文章を歌詞に最適な形にする

手法 4 原文から重要だと思われる単語を抽出し、単語を訳した後、日本語の歌詞を生成する

なお本章で示す実験では、日本語における撥音・促音のモーラ数を0とする。また、日本語の歌唱では1つの文字の母音を伸ばし複数の音符を歌いあげることが許容されていることから、日本語への翻訳によってモーラ数が原文より減る場合にはその訳をそのまま採用するものとする。

3.3.1 手法 1

手法 1 ではまず「原文の圧縮」を適用する。この処理の訳詞における合理性を検証するために、著者らはまず、歌曲の1フレーズにおける平均単語数を調査した。各フレーズの平均単語数が非常に小さければ、文圧縮を適用するに十分な単語数がないと判断できる。

著者らは英詞の歌曲を無作為に29曲選択し、人手でフレーズを分割した。結果として872のフレーズに分割され、全単語数は4,546であった。よって、1フレーズにおける単語数の平均値は5.386であった。以上の結果より、歌詞の1フレーズに相当する原文がとても短いため、文圧縮を適用しても単語数を十分に削減することは難しい。もしくは、無理に単語数を削減することで文意を損ない、日本語への翻訳が不能になることが懸念される。そのため、手法 1 は訳詞の手段として最適ではないとした。

3.3.2 手法2

手法2ではまず原文をそのまま翻訳し、得られた翻訳文からの操作によって訳詞を生成する。翻訳文からの処理過程で必要となる操作について検討するために著者らは、原文から得た翻訳文について考察した。

以下に、童謡「ロンドン橋」とミュージカル「Les Miserables」より一部抜粋した歌詞を示す。括弧内の数字は各フレーズのモーラ数を示す。翻訳には Bing 翻訳²を用いた。

London Bridge

London Bridge is falling down (7)
Falling down, falling down (6)
London Bridge is falling down (7)
My fair lady (4)

ロンドン橋が落ちる (8)
落ちる, 落ちる (6)
ロンドン橋が落ちる (8)
マイ・フェア・レディ (6)

Cafe Song

There's a grief that can't be spoken (8)
There's a pain goes on and on (8)
Empty chairs at empty tables (8)
Now my friends are dead and gone (7)

話すことができない悲しみがあります (19)
痛みは延々とがあります (12)
空のテーブル空の椅子 (12)
私の友人が死んでいっている今 (16)

ロンドン橋の結果について、1フレーズ目と3フレーズ目の格助詞「が」を削除すれば、各フレーズともモーラ数が原詞と同じになり、歌詞として成立することが伺える。さらに、第2フレーズはそのまま歌詞として利用することが可能であった。

その一方で、「Cafe Song」の原詞と訳詞ではモーラ数が大きく異なっている。また、体言止めと、ですます調が入り混じっており、歌詞の語尾に統一性がなく、この翻訳文をそのまま訳詞として利用するのは難しい。ここで、岩谷時子による「Cafe Song」の訳詞の例を示す。

²<http://www.bing.com/translator>

カフェ・ソング

話すことができない悲しみがありません。(19)

痛みは延々とあります。(12)

空のテーブル空の椅子(12)

私の友人が死んでいっている今(16)

言葉にならない(8)

痛みと悲しみ(8)

空の椅子とテーブル(10)

友はもういない(8)

「話すことができない悲しみがありません」を人手で訳した場合の歌詞は「言葉にならない」となっており、「話すことができない」のパラフレーズを求める必要性が伺える。また、第2フレーズの「痛みは延々とあります」を人手で訳した場合の歌詞は「痛みと悲しみ」であり、第1フレーズの「悲しみ」が第2フレーズへと移っており、入れ替えが発生していることがわかる。第4フレーズ目は「私」という主語が人手の翻訳では消えている。これは、歌っている本人が「私」であり、日本語では主語を省略しても十分に意味が通じるからであると考えられる。

これより、手法2による自動訳詞のための操作として以下が有効であることが示唆される。

1. 助詞を省略する
2. 翻訳文を元にパラフレーズを得る
3. フレーズの前後で単語を入れ替える
4. 一人称としての「私」という主語を省略する

(2)について、パラフレーズによってモーラ数を減らせる可能性があるのは明らかである。そのため、ここでは(1)(3)(4)について、(1)(4)を単語の省略についての課題とし、(3)を単語の倒置の課題として、検証を行う。

以下の実験では、機械翻訳機として、Bing 翻訳及び Google 翻訳³を用いて比較をした。この2つの課題において、調査に用いた楽曲数は23曲であり、全体のフレーズ数は652であった。

単語の省略

歌詞の原語を機械翻訳した結果（以後、翻訳文と呼ぶ）と、人手による訳詞文を用いて、単語を省略することで、翻訳文と訳詞文の類似度が上がるかを検証した。本実験では単純な調査として、人手による訳詞文を正解文として扱い、翻訳文から単語（格助詞および「私」）を省略したものを比較した。

³<https://translate.google.co.jp/>

表記ゆれを避けるため、訳詞は全て平仮名とし句読点および記号は含めないこととした。翻訳文には漢字の混じった文を用意し、形態素解析を適用した後で単語を省略し、平仮名へと変換して句読点を省略することで、平仮名の文章を生成した。その上で、翻訳文と訳詞文を1フレーズずつ比較し、類似度の平均値を求めた。本実験では形態素解析にはmecab⁴を用い、類似度の計算にはRatcliffらのPattern Matching: the Gestalt Approach[Ratcliff88]を利用した。

比較結果を表3.1に示す。類似度は最大を1.00とし、小数第4位を四捨五入した。

表 3.1: 単語の省略における類似度

比較内容	類似度	
	Bing 翻訳	Google 翻訳
翻訳文：訳詞文	0.408	0.398
格助詞を抜いた翻訳文：訳詞文	0.415	0.410
「私」を抜いた翻訳文：訳詞文	0.416	0.412
格助詞と「私」を抜いた翻訳文：訳詞文	0.424	0.425

格助詞と「私」のどちらか一方を省略した場合、省略しなかった場合よりも類似度は上がるが、格助詞と「私」の両方を省略したとき、どちらの翻訳機においても、最も類似度が上がることがわかった。以上の結果から、(1)(4)に関する単語の省略について、訳詞においてモーラ数を減らす手段として有効であることが示された。

単語の倒置

翻訳文の1フレーズに含まれる単語が、訳詞文の前後の1フレーズ及び同フレーズに含まれる割合を調査した。この時、メロディフレーズの繰り返し等による単語の重複は考慮せずに割合を求める。また、調査する単語は翻訳文、訳詞文ともにmecabの品詞において「名詞」「形容詞」「動詞」「副詞」に含まれるものを対象とし、その原型を比較する。この割合について、以下のように定義する。

$$\text{単語含有率} = \frac{\text{翻訳文・訳詞文ともに使用されていた合計単語数}}{\text{訳詞文の合計単語数}} \quad (3.1)$$

$$\text{フレーズ含有率} = \frac{\text{翻訳文・訳詞文に共通する単語が存在した合計フレーズ数}}{\text{訳詞文の合計フレーズ数}} \quad (3.2)$$

式3.1、式3.2ともに、最大値を1.00とし、小数第4位を四捨五入することで得る。その結果を表3.2に示す。

⁴<https://taku910.github.io/mecab/>

表 3.2: 単語の倒置における含有率

対象フレーズ	単語含有率		フレーズ含有率	
	Bing 翻訳	Google 翻訳	Bing 翻訳	Google 翻訳
前フレーズ	0.08	0.07	0.13	0.13
後フレーズ	0.07	0.08	0.12	0.14
同フレーズ	0.17	0.17	0.26	0.27

以上の結果より、同フレーズに翻訳文と訳詞に共通して含まれる単語が最も多いことがわかる。また、入れ替えにおいて、Bing 翻訳、Google 翻訳ともに前後フレーズ含有率に差異はほとんどなく、どちらに入れ替えが起きる傾向が強いかは、はっきりとは言えない結果となった。入れ替えが発生している割合についても、単語含有率、フレーズ含有率ともに非常に低い数値であり、単語の入れ替えは必須ではないことが伺える。

同フレーズにおいて、Bing 翻訳と比較して、数値が高く出ている Google 翻訳の単語含有率は約 0.26、フレーズ含有率は約 0.27 であった。この結果から、翻訳文に利用されている単語は、同フレーズであっても訳詞文にほぼ含まれていないということがわかる。以上により、翻訳文で得た単語をそのまま訳詞文の一部として利用することは難しいことが示唆される。

3.3.3 手法 3

手法 3 の訳詞手段としての妥当性を検証するために著者らは、井手上らの部分的機械翻訳手法 [Yamamoto13] を、実際に歌詞として使われている文に適用した。本実験ではミュージカル「Les Miserables」から歌詞を抽出し、対訳コーパスには Project Gutenberg や青空文庫などから作成されたものを利用した⁵。

表 3.3: 手法 3 による翻訳結果

表 3.4: Translation Result by Method Three.

原文 (モーラ数)	部分的表現
I dreamed that love would never die (8)	I dreamed
How can I live when we are parted (9)	How can I
This man could be my chance (6)	my chance
部分的翻訳結果 (モーラ数)	Bing 翻訳
うとうとした (6)	愛が死ぬことはない夢を見た
どうやって? (4)	私たちが別れているときにライブする方法
本当にそれがやってきた (11)	この男は、私のチャンスかもしれない

翻訳の結果の例を表 3.4 に示す。比較のため、原文をそのまま Bing 翻訳にかけた結果も併記している。

⁵<http://www2.nict.go.jp/astrec-att/member/mutiyama/align/index.html>

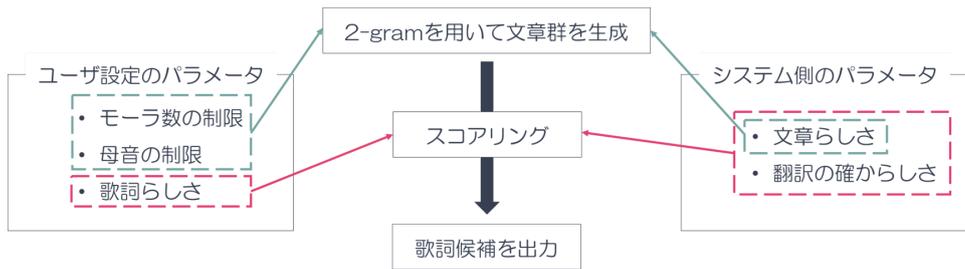


図 3.3: 歌詞候補群選出の手順.

この結果から、モーラ数が大幅に削減されている文章も見られるが、かえって原文よりもモーラ数が増えている例も存在することがわかる。また、翻訳の精度にもばらつきが見られる。その理由として、利用した対訳コーパスに部分的表現が一致する箇所が少なかったことなどが考えられる。加えて、対訳コーパス内に存在しない歌詞特有の表現を翻訳できないことから、訳詞結果が対訳コーパスに大きく依存してしまうことがわかる。

以上により著者らは、さまざまな文章表現が使用される歌詞の翻訳に手法3を適用するのは難しいと結論づけた。

3.3.4 手法4

手法4においては訳詞の際の課題点となるモーラ数および母音の制限に着目して2-gramを用いて生成した文章群から歌詞候補を選択する。歌詞候補を選択する過程において、文章らしさ、翻訳の確からしさ、生成された文章の歌詞らしさの3点を考慮する。

歌詞候補群

今回用いた歌詞候補群の生成手順を図3.3に示す。最初に、2-gramを用いて文章を生成することで文章群を得る。機械翻訳文から重要だと思われる単語を手手で1つ抜き出し、それを先頭とした、2-gramを取得する。

文章群を生成する際、制限を設けない場合、大量の文章が生成されてしまう。そのため、モーラ数の制限、母音の制限、文章らしさ、を用いて生成する文章群に条件を課す。取得した2-gramの後方の単語を先頭とする2-gramを更に取得し、指定されたモーラ数と一致するまで探索をする。このとき、探索の途中で形態素解析を行い、助詞が連続すると判断された文章は文章群の候補から外れる。また、特定の位置のモーラに対して母音の制限がある場合、条件を満たさない文章についても文章群の候補から外れる。

さらに、得られた文章群に、歌詞らしさ、文章らしさ、翻訳の確からしさをを用いてスコアを付けることによって、スコアの高い上位5位を歌詞候補群として選出する。撥音、促音は0モーラとして数え、モー

ラ数が減る場合は考慮せず、モーラ数が指定された数と一致するものだけを考察対象とする。また、形態素解析には MeCab を用いた。本実験において、コーパスは日本語版 Wikipedia コーパスを利用した⁶。

文章らしさ

2-gram の先頭の単語を w_{i-1} 、後方の単語を $w_i (i > 1)$ とし、生成文を s としたとき、2-gram 確率を $P(w_i|w_{i-1})$ 、生成文の生起確率を $P(s)$ として以下のように定義する。

$$P(w_i|w_{i-1}) = \frac{\text{利用コーパス内において } w_{i-1} \text{ の後に } w_i \text{ が続く頻度}}{\text{利用コーパス内における } w_{i-1} \text{ の出現頻度}} \quad (3.3)$$

$$s = \{w_1, w_2, w_3, \dots, w_n\} \quad (3.4)$$

$$P(s) = \prod_{i=2}^n P(w_i|w_{i-1}) \quad (3.5)$$

文章らしさは、モーラ数が増えるとともに、値が下がっていく。探索の途中において $P(s) > P_1$ を満たさなくなった文章については文章群の候補から外した。また、モーラ数と一致する文章において $P(s) > P_2$ となるものを文章群の候補とする。ここで、今回の実装では $P_1 = 1.14e-4$ 、 $P_2 = 0.11e-4$ とした。

翻訳の確からしさ

翻訳の確からしさには、原言語の機械翻訳文を利用する。単語のベクトル化には Word2Vec[Mikolov13] を用い、次元数を 200 とした。

機械翻訳文と生成文から、それぞれ名詞と動詞を抜き出し、cos 類似度を算出する。このとき、生成文中の単語に対して、機械翻訳文中の単語から cos 類似度が最も高くなる値を求める。生成文を s とし、生成文の集合を S としたとき、 s に含まれる対象となる単語を w 、 w に対して最も cos 類似度が高くなる単語を v 、機械翻訳文中の生成文の翻訳の確からしさを $T(s)$ としたとき、式 3.6 として定義する。

$$T(s) = \sum_{k=1}^n \max_{v_j \in S} (\cos(w_k, v_j)) \quad (s \in S) \quad (3.6)$$

生成された文章の歌詞らしさ

生成された文章が、必ずしも歌詞らしい文章であるとは限らない。そのため、ユーザが歌詞らしさをどれだけ重視するかを設定する必要がある。ここで、piapro⁷から生成したコーパスは、ある程度、歌詞

⁶<http://dumps.wikimedia.org/jawiki/latest/jawiki-latest-pages-articles.xml.bz2>

⁷<http://piapro.jp>

らしくない単語を除外できているという実験結果がある [Abe12]. そこで, piapro コーパス中に歌詞候補に含まれる単語が存在した場合は, 歌詞らしいと判断できると考えた.

歌詞らしさについて, 生成文を形態素解析し, 名詞, 動詞, 形容詞, 副詞を抜き出し調査対象の単語とする. そして, 生成文の単語が piapro コーパスに含まれる割合を調べる. 生成文を s とし, 生成文の歌詞らしさを $L(s)$ とした場合, 式 3.7 として定義する.

$$L(s) = \frac{\text{piapro コーパスに含まれていた単語の総数}}{s \text{ における調査対象の単語の総数}} \quad (3.7)$$

スコアリング

生成文を s とし, 文章らしさを $P(s)$, 翻訳の確からしさを $T(s)$, 歌詞らしさを $L(s)$ とし, 歌詞らしさにユーザが付与できる可変である重みを λ とし, 生成文に対して以下のように $Score(s)$ を求める. この $Score(s)$ が文章群中の生成文に対して最も高い 5 つを歌詞候補群とする.

$$Score(s) = P(s) + T(s) + \lambda L(s) (0.0 \leq \lambda \leq 1.0) \quad (3.8)$$

スコアリングによる評価と議論

表 3.5: 実験条件

グループ	指定モーラ数	先頭の単語	母音, モーラ位置
A-1	5	古い	指定なし
A-2	5	古い	(ア, 5)
A-3	6	古い	(ア, 5)
A-4	7	古い	(ア, 5)
A-5	8	古い	(ア, 5)
A-6	5	古い	(ア, 4), (ア, 5)

実験には, 原言語を「Tale as old as time」, 機械翻訳文を「時間として古い物語」, 実音数を「5」とする歌詞を使用した. 歌詞候補群生成のための条件を表 3.5 に示す. 手法 1 の結果より, 1 フレーズにおける単語数の平均値は 5.386 であることがわかっている. そのため, 音節などにより音数が増えることを想定し 5~8 の範囲で歌詞候補の探索を行った. また, 母音については, 0~2 音の範囲で指定することとした. この時, 先頭の単語および母音は人手で選択した. その実験結果を表 3.6 に示す.

今回の実験において, スコア上位 5 位に含まれる単語の全てが piapro コーパスに含まれていたため, 歌詞らしさの重みによる変動が無かった. そのため, 実験条件及び実験結果の歌詞らしさの重みは全て

表 3.6: 実験結果

グループ	生成文数	結果	スコア	グループ	生成分数	結果	スコア
A-1	982	古い時間	2.00005	A-4	354	古い話を	1.61809
		古い神話	1.57048			古い話が	1.61802
		古い時分	1.44857			古い話で	1.61795
		古い詩を	1.44435			古い話は	1.61795
		古い詩の	1.44433			古い話の	1.61793
A-2	100	古い詩は	1.44426	A-5	216	古い話数	1.99134
		古い詩が	1.44424			古い個体数	1.57231
		古い詩や	1.44422			古い時代ごと	1.55306
		古い歌	1.39155			古い時代のは	1.50141
		古い値が	1.38999			古い時代のを	1.50132
A-3	324	古い話	1.41879	A-6	37	古い球	1.26576
		古い舞台	1.27864			古い穴	1.23765
		古い場合	1.26686			古い花	1.22408
		古い葉先	1.26116			古いマヤ	1.21673
		古い語り	1.20745			古い朝	1.20135

$\lambda = 1.00$ として結果を示し、表記は省略する。実験結果をみると、概ね同じ単語を用いた意味の歌詞候補が選出されている。

A-1を見ると、「古い時間」という機械翻訳文に含まれる単語が2つ使用された文章が歌詞候補として1番高いスコアで選出されていることがわかる。しかしながら、これは、原言語の意味とは異なっている。一方で、A-3の「古い話」はモーラ数は1増えているが、原言語の意味を反映していると言える。この2つを比較したとき、スコアは「古い時間」の方が高く、翻訳の確からしさについて、スコアのつけかたが不十分であることがわかった。

A-4については、最後の助詞が変化しているだけで、全ての候補群の意味が同じとなっていた。これは、他のグループにおいても見られる問題である。助詞の選び方については、次のフレーズの歌詞によって変わるので、今後、複数のフレーズを考慮した歌詞候補の選出手法が必要になってくる。

A-5に着目すると「古い話数」とあるが、これは、形態素解析により読み方を「フルイハナシスウ」と判断されたためである。モーラ数は形態素解析の結果に依存してしまい、必ずしも正しい読みにはならない結果となった。

また、A-6を見ると、他と比較して文章群の生成数が大幅に減り、原言語とは関係の無い意味を持つ歌詞が候補としてあがっている。このことから、母音の指定数を増やすと、全く違った意味を持つ歌詞が候補としてあがることがわかった。これは、人手の訳詞においても懸念されていた点である。

これらの検証結果より、パラメータを調整し、文章群から歌詞候補を出すことについて以下のような課題点が挙げられる。

- 歌詞らしさは別の手段によってスコアをつける必要がある
- 翻訳の確からしさについて、精度を上げる必要がある
- 歌詞候補の最後の助詞について、次のフレーズを考慮した選出が必要である
- モーラ数は形態素解析の結果に依存する
- 原言語の意味を反映するためには、母音の限定は最低限に留める必要がある

歌詞らしさの指標については、テストセットパープレキシティを用いると改善する可能性がある。piaproのコーパスがある程度、歌詞らしくない単語を除外できているという結果が出ているため、piaproコーパスを用いて言語モデルを作成する。そして、生成された文章を入力文として、言語モデルが生成する確率を計算することで、パープレキシティの値を得られる。テストセットパープレキシティが低いほど、生成された文章は歌詞に近い文章だと言える。また、ユーザによる入力を想定した重み λ についても、今後は値が変化することによって、どのように結果が変わるかを精査する必要がある。

翻訳の確からしさについては、指標として多言語の埋め込みモデルを用いて、原文と生成文の文章の類似度を求めるという方法がある。訳詞においては意識が必要になるため、原文と生成文の類似度を直接求められる点で適切な指標であると考えられる。たとえば、Languageagnostic BERT Sentence Embedding(LaBSE)[Feng22] と Multilingual Embeddings from bidirectional Encoder Representations(ME5)[Liang22] を用いて \cos 類似度を使用して、文章の類似度が求められる。

課題点がある一方で「古い話」のように原言語の意味を反映した歌詞候補を得られることもわかった。パラメータの調整やスコアの付け方によっては、生成した文章群から歌詞候補を選択する手法は有用であると言えるだろう。

3.4 4つの手法の検討結果に対する考察

4つの手法について検討した結果として著者らは、手法2と手法4を用いた訳詞が現状でもっとも現実的であると判断した。

また、これらの検証結果より、翻訳文からモーラ数を減らすための手段として、以下のような操作が有効だと考えられる。

- 助詞の省略
- 翻訳文を元にパラフレーズを得る
- 一人称としての「私」という主語の省略

- 原文の単語から類義語を求め、翻訳を行う

3.5 訳詞アルゴリズムの提案

3.3章の結果より、次の2つの手法がモーラ数を減らして訳詞を行うために有効であることがわかった。

- 原文に変更を加えず日本語に翻訳し、その文章を歌詞に最適な形にする
- 原文から重要だと思われる単語を抽出し、単語を訳した後、日本語の歌詞を生成する

また、翻訳文からモーラ数を減らすための4つの操作から助詞の省略を適用し、モーラ数を減らすことを試みた。加えて、機械翻訳文の語尾が安定しない問題があるため、語尾の表現を整形し、体言止めを適用する。

訳詞アルゴリズムの手順は、(1) 原言語の翻訳、(2) 翻訳文の歌詞化、(3) 生成した歌詞のメロディへの当てはめ、の3つに大別することができる。

3.5.1 原言語の翻訳

入力された MusicXML の英語歌詞から、Microsoft Translator テキスト API⁸を利用して機械翻訳文を取得する。この時、ピリオド、カンマ、セミコロンを1フレーズとみなし、歌詞に付随するメロディの音符の数を英語歌詞の音数とする。以降、1フレーズに対する英語歌詞の音数をメロディ音数と呼ぶ。

翻訳はAPI依存であり、取得した機械翻訳文に、訳しきれなかった英語が残る場合がある。その場合は、アルファベットを削除する。

ここの処理では、1フレーズの機械翻訳文および、メロディ音数を取得する。

3.5.2 翻訳文の歌詞化

取得した機械翻訳文および、メロディ音数を利用して、歌詞の生成を行う。処理手順を図3.4に示す。1フレーズの機械翻訳文のモーラ数がメロディ音数を超える場合、次の4つの操作を逐次実行する。

Step1 敬語表現の削除

Step2 断定表現の削除

Step3 助詞の省略

⁸<https://www.microsoft.com/ja-jp/translator/translatorapi.aspx>

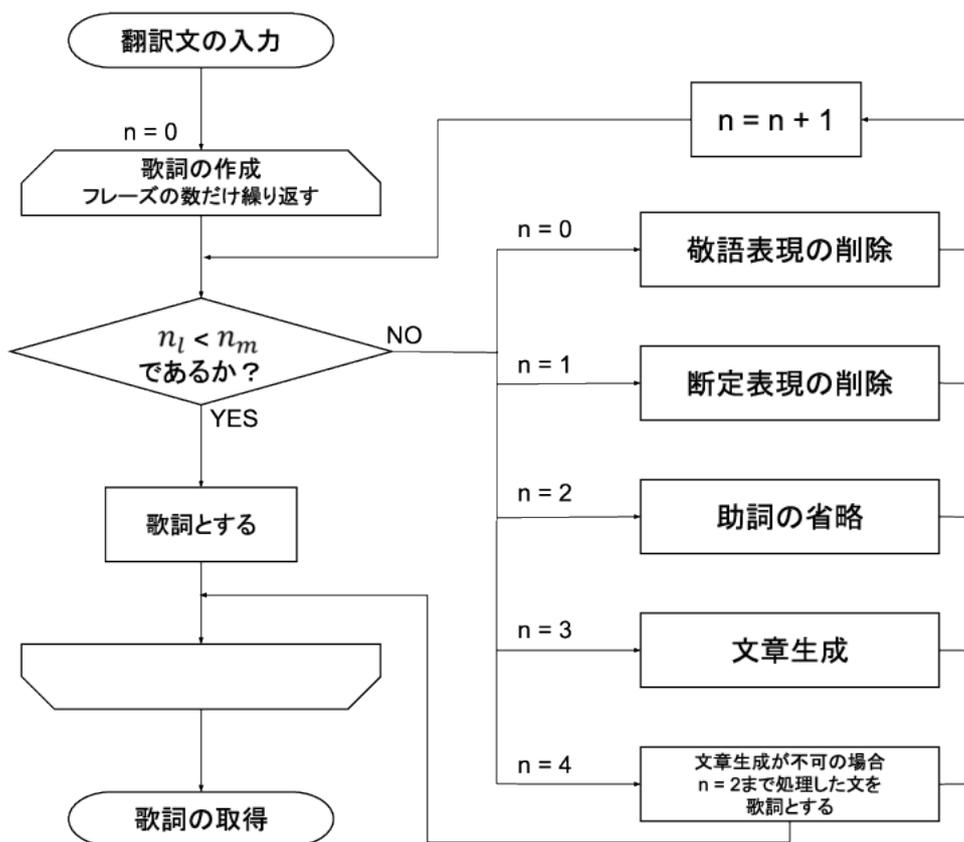


図 3.4: 歌詞化の処理手順.

Step4 文章生成

機械翻訳を利用した場合、多くの文章に敬語表現が使用されている。Step1 の敬語の削除において、モーラを減らす手段として、取得した機械翻訳文から敬語表現である「です」「ます」を削除する。Step2 の断定表現の削除において、文末表現が「だった」「である」「だ」というような断定の場合も、敬語表現と同様に削除する。Step3 の助詞の省略において、形態素解析の結果、助詞と判断された単語は翻訳文から削除する。

上記の3つの手順を踏んだ上で、メロディ音数を超える場合は、step4として、3.3.4節で示した手法を使い、1フレーズ中の tf-idf 値が最も高い単語を先頭の単語として与えることで、2-gram を用いてメロディ音数と同等のモーラ数を持つ歌詞を生成する。tf-idf (The term frequency-inverse document frequency) とは、文書に含まれる単語の重要度を評価するための統計値である。tf-idf 値を算出するための文書として、Project Gutenberg と青空文庫が作成したコーパス [Utiyama03] を使用した。一方で、モーラ数と与えられた単語によっては歌詞の生成ができない場合もある。その場合は、助詞の省略まで行った文章を歌詞として、メロディ音数を超えることを許容する。

この処理を、全てのフレーズに対して繰り返し、鼻濁音のモーラ数は0とした。この時、撥音および

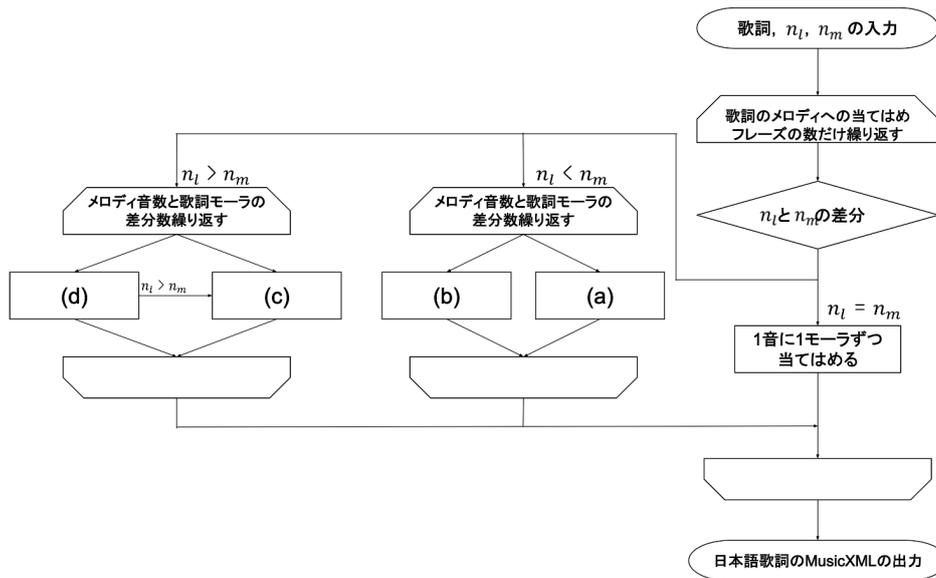


図 3.5: 生成した歌詞のメロディへの当てはめの処理手順.

促音は 0 モーラとして数える。また、形態素解析には MeCab を用いた。

3.5.3 生成した歌詞のメロディへの当てはめ

生成された歌詞のモーラ数 n_l とメロディ音数 n_m の関係を次の 3 種類に定義する。(1) $n_l = n_m$, (2) $n_l < n_m$, (3) $n_l > n_m$ とする。 n_l と n_m が等しくても、メロディと歌詞の区切りが一致しないことが起こり得るが、簡単のため今回のアルゴリズムでは考慮しない。代わりに、この問題が残ることを許容し、ユーザーが機械翻訳された歌詞を編集し、区切りの不一致を改善することを期待する。生成した歌詞をメロディへ当てはめるための処理手順を図 3.5 に示す。 n_l と n_m が (1) を満たす場合、メロディを修正する必要はない。(2) または (3) を満たす場合、メロディと歌詞のマッチングに使用するそれぞれ 2 つのアルゴリズムのうちの 1 つがランダムに適用される。

日本語の歌唱では通常、一文字の母音を伸ばして複数の音を歌うことができる。ここでは、音の長さの関数として $dur()$ を定義する。(2) $n_l < n_m$ では、フレーズの先頭から i 番目の音符の音価が $dur(i) < dur(i+1)$ を満たすとき、図 3.5-(a) に示すように、フレーズ中の i 番目の音符に音引き「ー」を挿入することにより、歌詞のモーラ数を増加させる。先述のアルゴリズムは、すべての音符の長さが等しいとき、歌詞の最後に音引きを挿入する。2 つ目のアルゴリズムは、フレーズ内で最も短い音符を、その音符の 1 つ前の音符に統合する (図 3.5-(b))。これが不可能な場合、アルゴリズムは最も短い音符を次の音符に統合する。

(3) $n_l > n_m$ では、フレーズの先頭に最も近い最も音の長い音符の長さが、2 つの音符に分割される。付点音符の場合は、2 : 1 の長さの音符に分割する (図 3.5-(c))。2 つ目のアルゴリズムは、あらかじめ最

Generated Lyrics
 Lon - don ba - shi o - chi - ru o - chi - ru o - chi - ru

Manually Translated Lyrics
 ろん どんばし おちる おちる おちる

General English Translation
 London bridge is falling down falling down falling down

5
 Lon - don ba - shi o - chi - ru ma - i fe - a re - di
 ろん どんばし おちる まいふえ あれでい
 London bridge is falling down my fair lady

図 3.6: London Bridge の訳詞結果.

Generated Lyrics
 Ku - ri - i - ki no shi - i - ta de a - na - a - ta to wa - ta - shi shi - a - wa - se

Manually Translated Lyrics
 おおきなくりのきのしたで あなたとわたし なかよく

General English Translation
 Under a chestnut tree You and I A happy

6
 ta - bi - na - ni Ku - ri - i - ki no shi - i - ta de
 あそびましょ おおきなくりのきのしたで
 trip what Under a chestnut tree

図 3.7: Under the Spreading Chestnut Tree の訳詞結果.

小の音符の長さを定義しておき、定義した音符の長さより短くならないように音符を分割するものである (図 3.5-(d)). ここでは 8 分音符を最小音符長としている. すべての音符を最小音符長で分割しても, n_l が n_m より大きくなれば, さらに (c) のアルゴリズムが適用される.

3.6 訳詞アルゴリズムのモーラ誤差率による評価と結果

表 3.7: 訳詞アルゴリズムを適用した曲の RMSE.

曲名	RMSE
London Bridge	0.89
Under the Spreading Chestnut Tree	0.87
The Itsy Bitsy Spider	1.50
Scarborough Fair	5.27

専門家の訳詞と本論文で提案した訳詞アルゴリズムのモーラ数増減を比較することで, 訳詞アルゴリ

ズムがどの程度モーラ数の増減を抑えられたかを評価する。

はじめに、専門家が日本語に訳した訳詞において、原言語と訳詞のモーラ数の誤差率を調査した。櫻井らは歌唱可能な歌詞かどうかを測るモーラ数の誤差率を RMSE として算出した [Sakurai23]。本論文でも、RMSE を評価指標として使用する。

n を入力フレーズ数、 a_i を原言語歌詞のモーラ数、 p_i を訳詞された歌詞のモーラ数とする。

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (a_i - p_i)^2}{n}} \quad (3.9)$$

プロが翻訳した楽曲の訳詞から 329 フレーズを抽出して RMSE を算出した。3 つの演目よりそれぞれ、「マリツァ伯爵令嬢」から 111 フレーズ、「美女と野獣」から 106 フレーズ、「Les Miserables」から 112 フレーズである。その結果、 $RMSE = 1.09$ となった。

「London Bridge」「Under the Spreading Chestnut Tree」「The Itsy Bitsy Spider」「Scarborough Fair」の童謡を中心とした 4 曲の歌曲に訳詞アルゴリズムを適用し、RMSE を求めた。童謡を選択したのは、比較的平易な言葉で歌詞が書かれているため、英語から日本語へと翻訳する段階での誤訳が少ないと考えたためである。曲全体では $RMSE = 2.73$ となり、プロの訳詞と比較するとモーラ数の誤差が大きくなる結果となった。それぞれの歌曲の結果を表 3.7 に示す。「Scarborough Fair」は特に大きな差を生む結果となった。これは、他の曲に比べて歌詞の意味が複雑であることが原因だと考えられる。「The Itsy Bitsy Spider」は、数値だけに注目すると専門家の訳詞の RMSE に近い値だが、訳詞された歌詞は、ほとんどのフレーズで意味を成していなかった。一方で「London Bridge」と「Under the Spreading Chestnut Tree」においては良い結果を得ることができた。

図 3.6 と図 3.7 に訳詞の結果を示す。「Generated Lyrics」は、訳詞アルゴリズムによって生成された歌詞を表し、「Manually Translated Lyrics」は日本で翻訳されたポピュラーな訳詞を表す。「General English Translation」は歌詞を表しているのではなく、「Generated Lyrics」を機械翻訳した結果である。

どちらの結果も人手の訳詞と同じ歌詞部分が存在した。しかしながら、「Under the Spreading Chestnut Tree」においては意味を理解するのが難しい訳詞部分もある。音楽的な部分では「London Bridge」の最後のフレーズである「マイフェアレディ」という歌詞が、非常に歌いにくい歌詞となっている。英語では「my fair」と、2 音節になっている部分が日本語に訳したことで、「ま・い・ふえ・あ」のように 4 音節になってしまったためである。

これらの結果から、訳詞アルゴリズムは、簡単な同様であれば、人での訳詞と同じ歌詞が出力されることが分かった。しかしながら、モーラ誤差率が低くとも、歌詞の意味が通じない部分も存在する。本論文ではユーザの編集が入ることを前提としているため、訳詞の際に最低限の意味を担保することを目

指す必要がある。

3.7 むすび

本章では、外国語の歌曲を日本語に訳詞する場合の課題点を挙げ、訳詞を行う手法についての検討および考察について述べた。その結果、手法 2(3.3.2節) および手法 4(3.3.4節) が有効であるという結論を得た。さらに、3.5節で訳詞のアルゴリズムの提案を行い、機械翻訳文に対する処理手順や歌詞とメロディの当てはめ方をルール化した。3.6節でモーラ増減による評価を行った。その結果、童謡のような簡単な曲であれば人手の訳詞と同じ部分が存在することが認められた。一方で、複雑な曲ではモーラ数が極端に増えたり、モーラ数は変わらないが意味が通らない訳詞が出力されることがある。

本論文では、対話型の訳詞補助システムの構築を目標としている。そのため、今後はより歌詞の意味を担保できるように改善していきたい。

第4章 訳詞インタフェースの予備実装

4.1 はじめに

本章では、訳詞の作業を行うために適しているインタフェースを検討するために、各機能に対して2つの異なる実装を行い、任意に切り替えられるシステムの開発についての説明を行う。さらに、19名の被験者に対してシステムを使用して訳詞の作業を行ってもらった結果についても述べる。

4.2 システムの予備実装

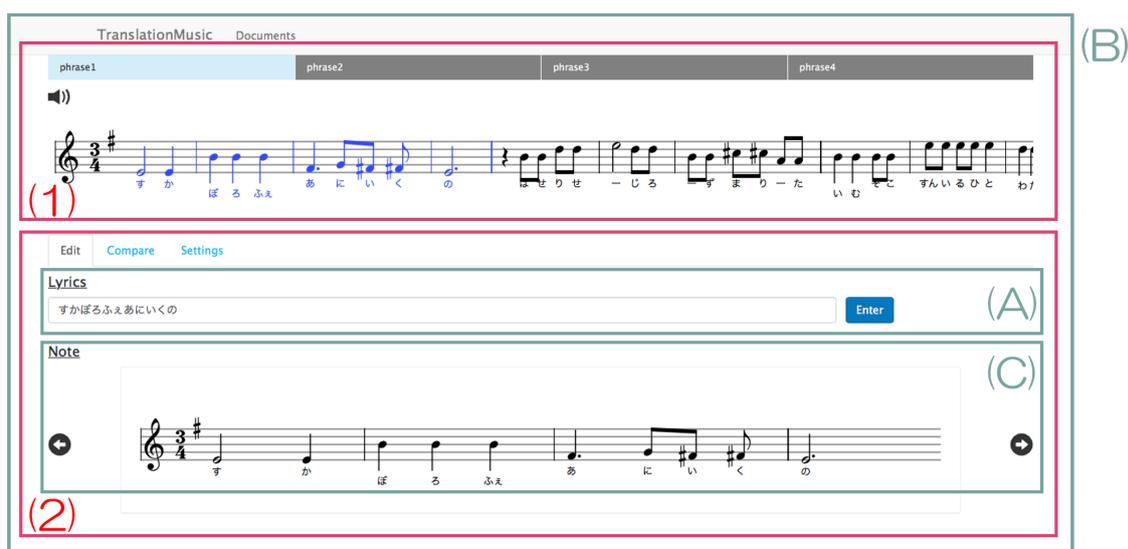


図 4.1: インタフェース選定用のシステム.

表 4.1: 各機能のモジュール内容

	(A) 訳詞候補	(B) 歌詞編集	(C) 音符編集
0	複数の訳詞候補から選択	初めに原語歌詞のみを見て複数の訳詞候補から選択	自動で編集
1	訳詞候補は1つのみ表示	訳詞が出力された譜面から任意で編集	候補から選択

本論文では、訳詞の作業環境としてシンプルな操作で編集が可能なインタフェースの提供を目標としている。しかしながら、操作がクリックやドラッグだけであったとしても、ユーザへの提示方法は多くの種類が考えられる。そのため、訳詞のためのシステムとしてどのようなインタフェースが適している

のかを検討するための実装を行った。本実装では各機能をモジュール化し、2種類の実装を任意に切り替えられる設計にした。図 4.1に、それぞれの機能を切り替え、構成要素の組み合わせの変更を可能にしたシステムのインタフェースを示す。実装したシステムの画面の機能は (1) 譜面表示部分、(2) タブ機能部分の2つの構成要素に大別される。

譜面表示部分はインタフェースの上部に表示される。譜面表示部分の上部には楽曲中のどのフレーズを編集しているかを示すバーがあり、対象フレーズを青色で表示する。ユーザは譜面を左右にスクロールすることでフレーズを選択できる。また、譜面左上のスピーカーアイコンをクリックすると、編集中の楽曲を試聴できる。本実装の視聴機能では、Sinsy[Oura10]を使用し、編集中の楽曲の全体の音源がクリック時に動的に生成される。Sinsyは、隠れマルコフモデルとディープニューラルネットワーク(DNN)を用いて、MusicXMLで記述された楽譜をもとに歌声を生成する。

タブ機能部分では、Edit、Compare、Settingsの3種類のタブが用意してある。Editタブでは歌詞および音符の編集ができる。Compareタブでは英語歌詞と日本語歌詞を比較できる。Settingsタブではアプリケーションのヘルプ画面表示や、譜面の編集を終了するためのボタンがある。

今回実装したシステムでは(A)訳詞候補、(B)歌詞編集、(C)音符編集の3つの機能を、それぞれについて2種類の実装を任意に切り替えられる設計にした(表 4.1)。各機能には2種類の実装があり、それぞれ0または1と表記し、例えば、(A)訳詞候補の機能はA0またはA1とする。8通りの機能の組み合わせのうち、A1とB0の組み合わせでは訳詞候補が1つしかない場合(A1)、フレーズごとの候補は一意になるため、複数の訳詞候補から選択する(B0)という操作に矛盾する。そこで、予備実装では、この2つの組み合わせを除外し他6通りのいずれかがランダムに適用される仕様とした。続いて各モジュールの機能の詳細を述べる。

4.2.1 訳詞候補



図 4.2: A0 と A1 のインタフェース。

訳詞候補機能には図 4.2に示すように、プルダウンメニューとテキストフィールドの2種類がある。A0では4つの訳詞候補を提示し、プルダウン形式の表示によって選択させる。予備実装では訳詞アルゴリ

ズムを用いて出力した候補を第1候補とし、第2候補から第4候補には人手で作成した訳詞を用いている。訳詞アルゴリズムのうち歌詞生成には多くの時間がかかるため(ステップ4)、予備実装ではあらかじめ生成した訳詞を表示する。すべての訳詞候補において、英語歌詞の機械翻訳文中で最も tf-idf 値が高かった単語を訳詞の先頭に使用している。テキストフォームにはプルダウンから選択した訳詞が表示され、ユーザが自由に編集できる。

A1では第1候補の訳詞のみがテキストフォームに表示される。なお现阶段では、テキストフォームへの入力はひらがなに限定している。

4.2.2 歌詞編集

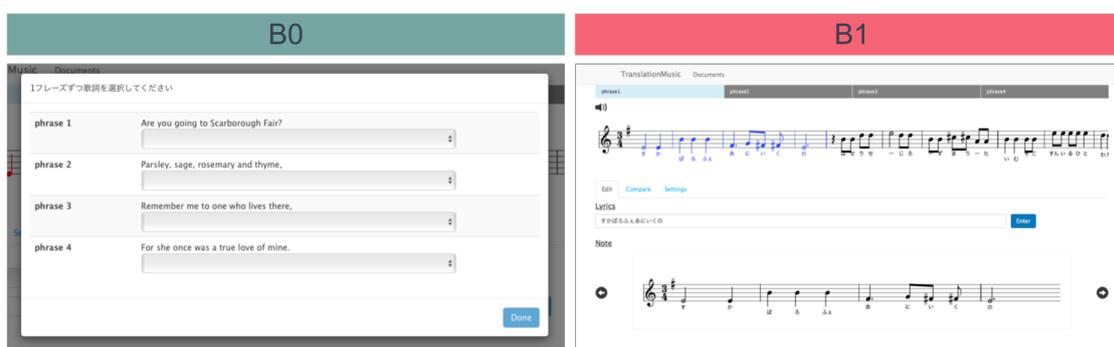


図 4.3: B0 と B1 のインタフェース。

B0ではまず図4.3-(B0)に示すようなモーダルが表示される。ユーザは譜面を確認する前に、原語歌詞のみを見て、1フレーズずつプルダウンの訳詞候補から訳詞を選択する。すべてのフレーズの訳詞を選択すると、右下のボタンが押せるようになり、ボタンを押すとモーダルが消え、B1と同様の編集画面が表示される。このとき、B0に対する組み合わせがA0の場合は、モーダル上で選択した訳詞候補が編集画面でも初期状態で表示される。また、テキストフォームにはA0、A1ともにモーダル上で選択した内容が表示されている。

B1では、システムにアクセスすると編集画面が表示され、ユーザは譜面にあらかじめ訳詞が入っている状態から、編集を開始する。

4.2.3 音符編集

音符編集は歌詞のモーラ数とメロディの音符の数が等しくない場合に必要になる作業である。ここでは3.5.3章で示したアルゴリズムを使用する。



図 4.4: C0 と C1 のインターフェース.

C0 を適用する場合、音符は自動的に分割または結合され、歌詞のモーラ数と音符の数が等しくなるため、ユーザによる操作は必要ない。今回の実装では、歌詞とメロディのマッチングに2つのアルゴリズムのうち1つをランダムに適用する。

C1 ではタブ機能部分の Edit 下部に譜割りをを選択するモジュールが表示される (図 4.4-(C1))。中央の譜面には歌詞とメロディの譜割り候補が表示される。 $n_l > n_m$ または $n_l < n_m$ のどちらの場合でも、最大で2種類の候補が表示される。ここで左右の矢印のボタンをクリックすることで、候補の表示が切り替わる。中央の譜面を押すと、その譜割りが選択され、画面上部の譜面表示部分に表示される。譜割りが未選択であれば譜面表示部分から歌詞が消え、フレーズの表示が赤色に切り替わる。

4.3 インタフェース構成のための予備実験

表 4.2: 被験者の分類

被験者の分類	組み合わせ	被験者数
A0 group	A0 + B1 + C0	6
	A0 + B1 + C1	
A1 group	A1 + B1 + C0	6
	A1 + B1 + C1	
B0 group	A0 + B0 + C0	7
	A0 + B0 + C1	
B1 group	A0 + B1 + C0	6
	A0 + B1 + C1	
C0 group	A0 + B0 + C0	9
	A0 + B1 + C0	
	A1 + B1 + C0	
C1 group	A0 + B0 + C1	10
	A0 + B1 + C1	
	A1 + B1 + C1	

インタフェースの構成要素の最適な組み合わせを決定するための予備実験を行った。本システムをサーバ上で稼働し、限定公開することで実験環境を用意した。ユーザがシステムにアクセスすると、システ

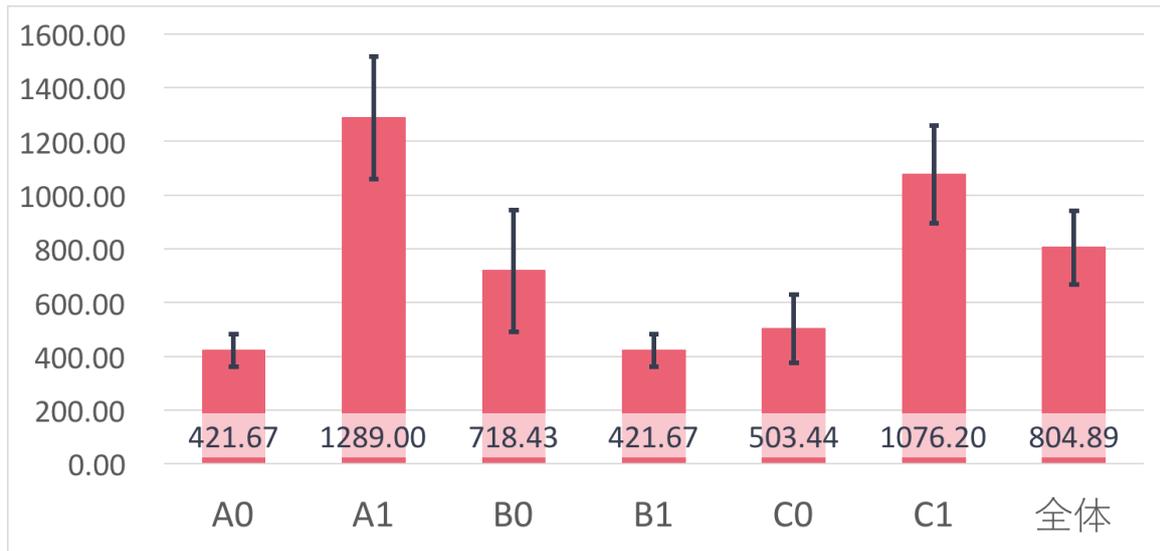


図 4.5: 各モジュールの平均作業時間と標準誤差.

ムの説明が表示されて訳詞の作業に入り、作業を完了するとアンケート入力画面に切り替わる、という流れでデータを収集した。訳詞作業の工程において、システム上でユーザがクリックした場所と回数および作業時間、表示されたモジュールの組み合わせを記録した。実験終了後、モジュールの組み合わせごとに被験者を分類し、それを母集団としてモジュール別の記録値の平均値および標準誤差を算出した。

作業後のアンケートではシステムの使いやすさ (1: とても使いにくい, 5: とても使いやすい), 完成した曲への満足度 (1: とても不満, 5: とても満足) を 5 段階のリッカート尺度で評価してもらった。加えて、システムの良かった点、悪かった点を自由記述で記入してもらった。

今回の実験では訳詞を行う曲として、フレーズ数が 4 つである「Scarborough Fair」を選択した。被験者は男性 9 人、女性 10 人、合計 19 人、23 歳～37 歳の平均年齢は 27 歳であった。最適なインターフェースを選択するために、Web アプリケーションに精通している被験者を集めた。そのため、被験者全員が Web エンジニアまたはデザイナーである。また、参加者のうち 10 名は楽譜を読むことができた。被験者の分類内容と被験者数を表 4.2 に示す。結果を集計、比較するために被験者をモジュールの組み合わせによって分類した。また、全ての数値を四捨五入により小数第 2 位まで示してある。

図 4.5 に各モジュールの平均作業時間と標準誤差を示す。A0 と B1 は 421.67 秒となり、他のモジュールより明らかに作業時間が短い。これは訳詞候補を提示することで作業時間を短縮できることを示唆している。続いてモジュールごとに結果を比較する。

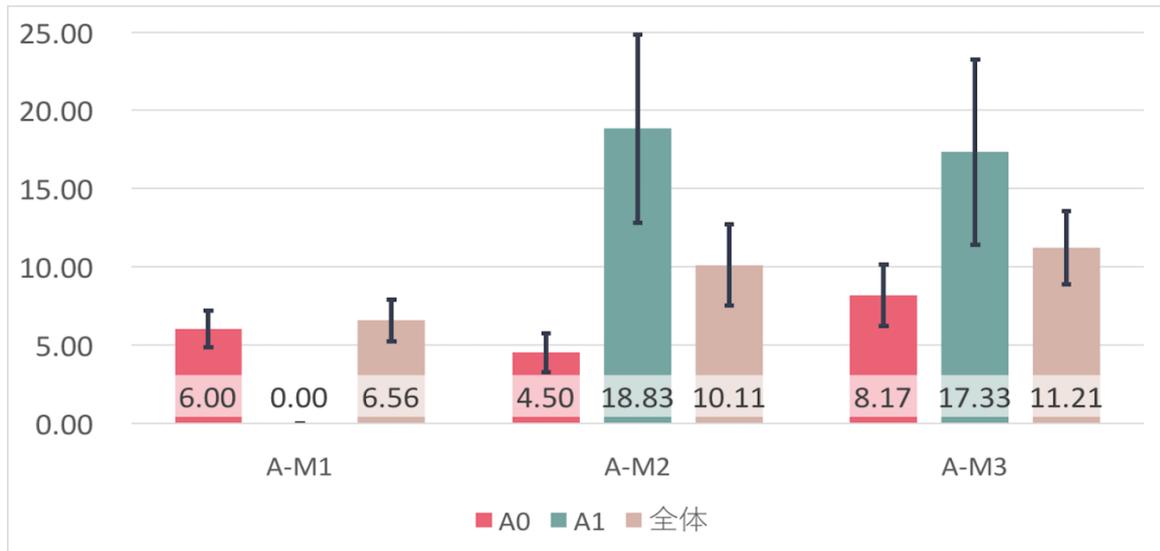


図 4.6: A0 と A1 の比較結果.

4.3.1 訳詞候補の結果

A0 と A1 について、クリック平均回数および標準誤差の比較結果を図 4.6 に示す。グラフはそれぞれ以下の内容を表している。

A-M1 プルダウンから訳詞候補を選択

A-M2 テキストフォームで歌詞を編集し決定した

A-M3 Compare タブを押した

A-M1 において、A1 が 0.00 となっているのは、モジュールとしてプルダウンが表示されていなかったことを表している。A0 において、プルダウンがクリックされた平均回数は 6.00 回だった。

A-M2 では、A0 のテキストフォームでの平均編集回数は 4.50 回となり、A1 の 18.83 回と比較すると 4 倍以上の差が見られた。全体と比較しても、訳詞候補が選択できる A0 は、ユーザーがテキストフォームから歌詞を編集する回数は少なかった。

A-M3 についても、A0 は 8.17 回、A1 は 17.33 回となり、A0 では英語と日本語の歌詞を比較する回数が少ないことがわかった。

また、図 4.5 より、A0 の作業時間は 421.67 秒であったのに対し、A1 の作業時間は 1,289.00 秒であった。

これらの結果から、A0 を使用した場合、ユーザーの編集回数や作業時間が短くなることが示唆された。したがって、訳詞候補ではプルダウンを用いて複数の候補から選択する A0 を適用すべきだと結論づけた。

4.3.2 歌詞編集の結果

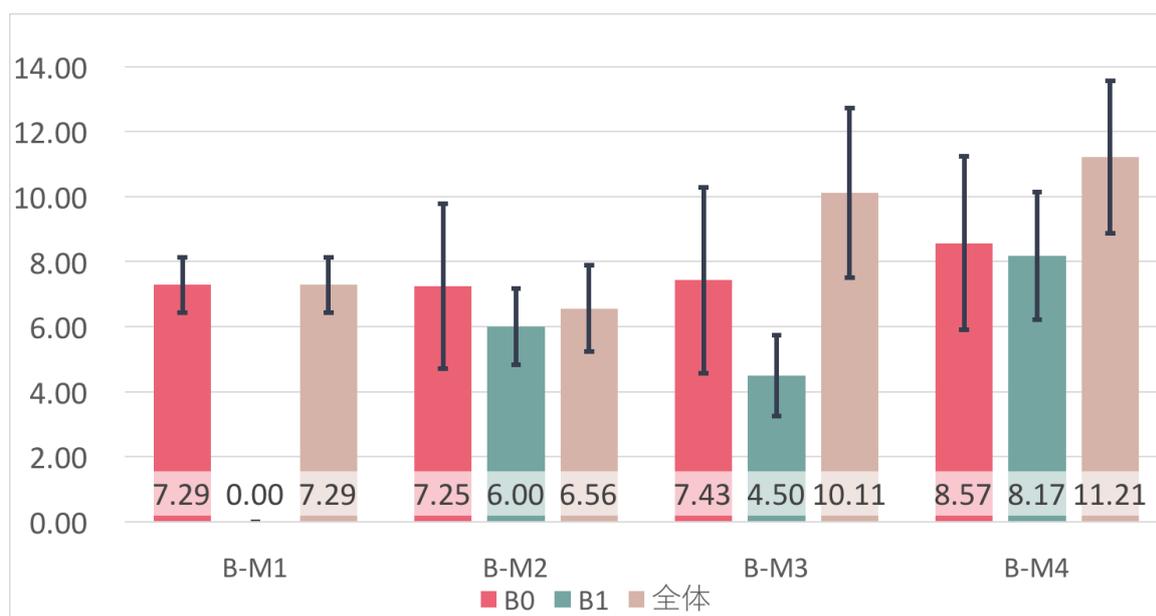


図 4.7: B0 と B1 の比較結果.

B0 と B1 について、クリック平均回数および標準誤差の比較を図 4.7に示す。グラフはそれぞれ、以下の内容を表している。

B-M1 モーダル上のプルダウンから訳詞候補を選択

B-M2 プルダウンから訳詞候補を選択

B-M3 テキストフォームで歌詞を編集し決定した

B-M4 Compare タブを押した

B-M1 において、B1 が 0.00 となっているのは、システムへのアクセス初期画面でモーダルが表示されないことを示している。B0 でモーダルを表示した際に、訳詞候補のプルダウンがクリックされた平均回数は 7.29 回だった。

B-M2 では、B0 について、最初にモーダル上のプルダウンから選択したにも関わらず、その後の編集画面にてプルダウンをクリックした回数は 7.25 回となっており、B1 の 6.00 回を上回った。

B-M3 にて、テキストフォームから歌詞が編集された回数についても、B0 が 7.43 回、B1 が 4.50 回と B1 の方が作業回数は少なかった。A の比較結果も踏まえると、A0 と B1 は他のモジュールと比較して作業回数が少ない傾向にあることがわかった。

B-M4 に注目すると、B0 と B1 はどちらも訳詞候補が表示されるが、全体と比較しても訳詞候補がある場合は、Compare タブを押した回数について、クリックされる回数が低いことがわかった。

また、図 4.5より、作業時間については B0 が 718.43 秒、B1 が 421.67 秒であった。

モーダル上で歌詞の訳詞候補を選択した場合でも、クリック数、編集回数、作業時間においても B1 の方が少ない回数や時間である。加えて、画面遷移の回数を減らすという点では、B1 の方が B0 よりも優れていると考えられる。そのため、今回は B1 をインターフェースとして採用することにした。

4.3.3 音符編集の結果

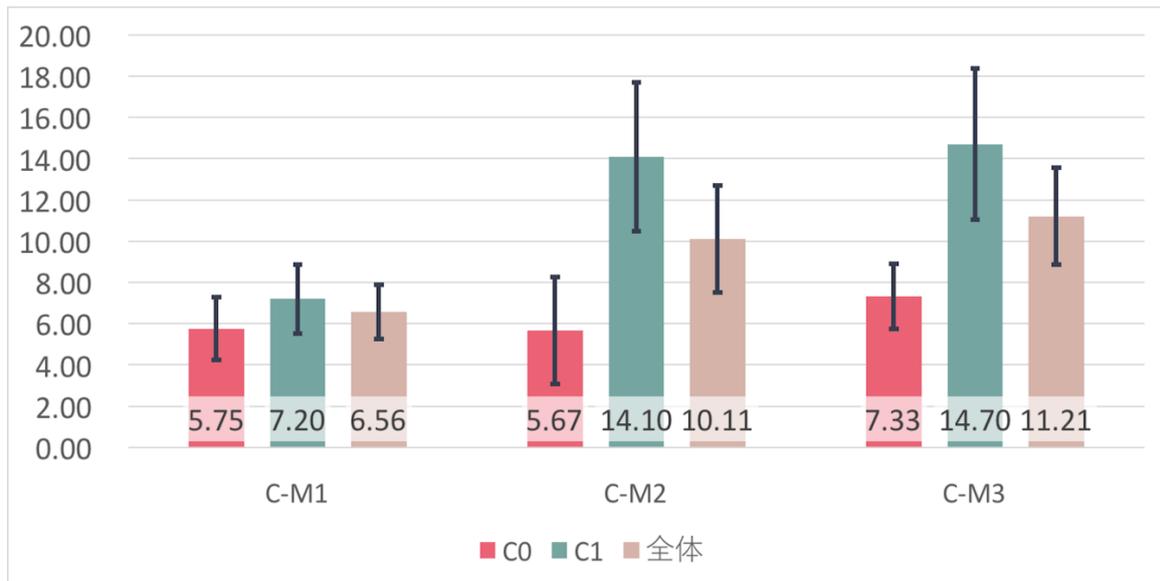


図 4.8: C0 と C1 の比較結果.

C0 と C1 について、クリック平均回数および標準誤差の比較を図 4.8に示す。グラフはそれぞれ、以下の内容を表している。

C-M1 プルダウンから訳詞候補を選択

C-M2 テキストフォームで歌詞を編集し決定した

C-M3 Compare タブを押した

C-M1、C-M2、C-M3 のすべての項目において、C0 の方が作業回数が少ない結果となった。全体と比較しても、C1 は作業回数が増える傾向が伺えた。

また、図 4.5より、作業時間については C0 が 503.44 秒、C1 グループが 1076.20 秒であった。

これらの結果より、C0 を使用した場合の方が作業回数が少なく、作業時間も C0 の方が C1 よりも短かったことが明らかになった。以上より、音符は自動的に編集されるべきだと結論づけた。

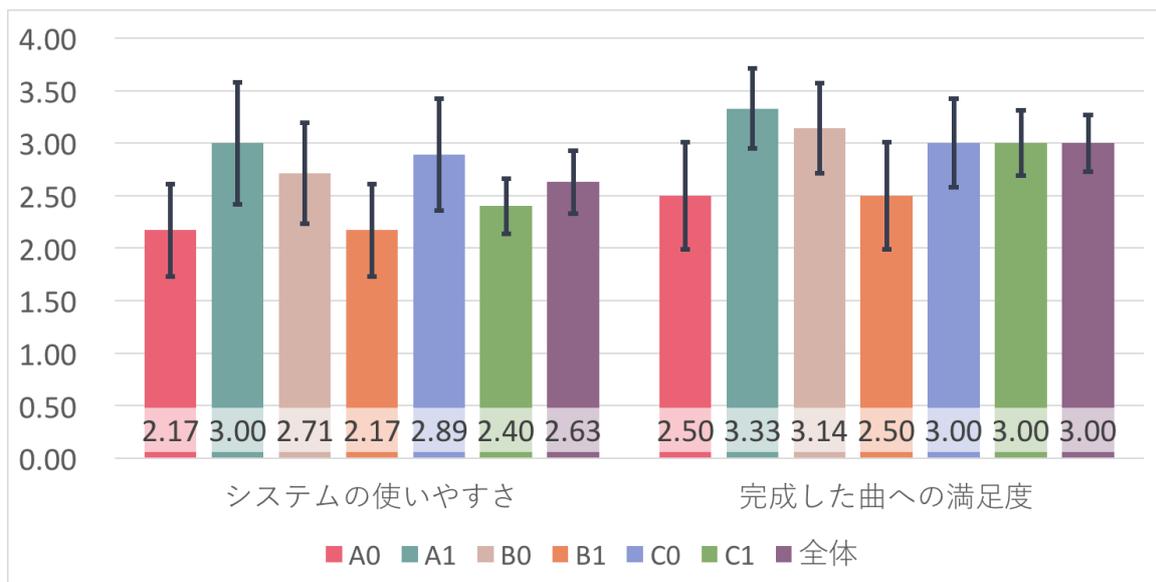


図 4.9: システムに関するアンケート結果.

4.3.4 アンケート結果

作業後のアンケートより、被験者から得た意見を以下に述べる。

システムの良かった点

- 訳詞を候補から選択できる (3 人)
- 歌詞の編集結果を歌唱として試聴できる (3 人)
- 全体的に見た目がわかりやすく操作しやすい (2 人)
- 横スクロールでフレーズが切り替わる (3 人)
- 音符編集を決定すると譜面の色が変わるので、作業が完了したのがわかりやすい

システムの悪かった点

- 譜面上部のフレーズを示すバーをクリックして移動したい (3 人)
- フレーズごとに試聴したい (3 人)
- 最初に何をすればいいかわからなかった (2 人)
- 編集する部分に英語を表示してほしい
- 音符編集の選択肢がいくつあるかわからない
- メロディに歌詞をいい感じに割り振る方法がわからなかった
- 各タスクを区切って示してほしい

アンケートの結果を図 4.9 に示す。A0 と A1 を比較すると、A0 の方が操作回数は少なかったにも関わ

らず、システムの使いやすさ、完成した曲に対する満足度は A1 が全て上回っていた。プルダウンから歌詞を選択できる点が良かったという意見が多かったことから、これらの値が低い理由として、悪かった点にも列挙されているとおり、他のインターフェースの不十分な点が影響していると考えられる。B0 と B1 では、作業回数の多かった B0 が、どちらにおいても B1 を上回る結果となった。A においても作業回数の多い A1 の満足度が高かったことを考えると、大きな作業回数と作業時間を経ることが完成した曲への満足度に寄与している可能性がある。

C0 と C1 を比較すると、システムの使いやすさは C0 が C1 を上回ったが、完成した曲に対する満足度は変わらなかった。

4.4 考察

実験の結果から以下の点がわかった。

1. 訳詞候補があると作業回数が少なく、作業時間も短縮される
2. 訳詞候補があると英語と日本語の比較作業が減る
3. 最初にモーダルを表示した方が使いやすいと感じるユーザーが多い
4. 歌詞の編集での作業回数が多いほうが満足度が高くなる傾向がある
5. 音符編集は自動編集の方が作業回数と作業時間が少なく、システムも使いやすく、満足度も変わらない

3. について、説明の直後に編集画面を出した場合、最初に何をすればいいかわからなかったという意見があった。そのため、最初にモーダルを表示し、強制的にプルダウンからユーザーに訳詞を選択させることで、作業内容が明確化したため、システムの使いやすさにポイントが入ったと推測する。これは、ユーザーのシステムに対する習熟度で評価が変わる可能性がある。そこで、クリック平均回数と作業時間を考慮すると、モーダルは表示せずに譜面から気に入らない部分を編集してもらう方が作業効率が良い。

以上の考察結果から著者らは、訳詞システムの開発は以下のような方向性が妥当であると判断した。

- 訳詞はプルダウンによる複数候補を提示する
- 選択された訳詞が出力された譜面から、必要であればユーザーが訳詞を編集する
- 音符は訳詞に合わせて自動編集する

4.5 むすび

本章では、訳詞システムのインタフェースを決定するために2種類の機能を任意に切り替えられるシステムの予備実装を行い、19名の被験者による予備実験の結果について述べた。クリック回数、作業時間、被験者から得た意見の結果から、「訳詞候補」「歌詞編集」「音符編集」の機能で使用するインタフェースを以下のように決定した。

訳詞候補 訳詞はプルダウンによる複数候補を提示する

歌詞編集 選択された訳詞が出力された譜面から、必要であればユーザが訳詞を編集する

音符編集 音符は訳詞に合わせて自動編集する

第5章 訳詞インタフェースの実装と評価

5.1 はじめに

本章では、4章で得た結果をもとに、5.2節でシステム全体の構成について説明し、5.3節で改良したシステムのインタフェースについて述べる。5.4節で、手書きでの訳詞とシステムを用いた訳詞の比較実験の内容と結果を説明する。さらに、比較実験で得た訳詞の印象についてアンケートによる第三者評価(5.5節)を実施した内容と結果を述べる。5.6節ではRMSEとcos類似度を用いて、比較実験で得た訳詞、さらに、GPT 3.5によって出力した訳詞に対して定量的な評価を行った結果と考察を述べる。

5.2 システム概要

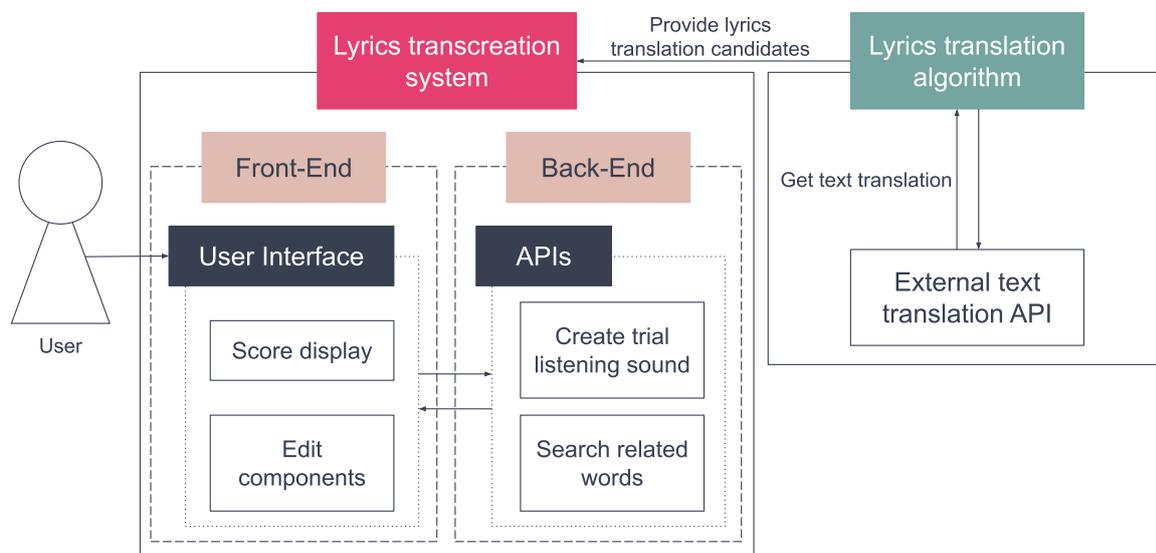


図 5.1: システム概要図.

本章で実装したシステムの構成を図 5.1に示す。改良したシステムの実装は、訳詞アルゴリズムと訳詞システムの2つの構成要素に分けることができる。

訳詞アルゴリズム部分では、外部のAPI (Application Programming Interface) を用いて原語の歌詞を翻訳し、翻訳されたテキストを歌詞の形式に加工する。

訳詞システムは、バックエンドとフロントエンドの2つのサーバを保持している。Sinsy と関連語検

索のためのモデルはバックエンド上に API として実装されている。フロントエンドでは、曲をフレーズごとに編集するためのインタフェースを web アプリケーションとして提供している。

4章で得た結果を踏まえ、インタフェースの改善を図った。編集機能ごとのタブを撤廃し、全機能を一つにまとめることで、情報の提示をより分かりやすくした。訳詞を編集する際に原言語の歌詞をテキストフィールドの上部に表示することで、ユーザーが自身の訳詞と比較しやすくなった。また、関連語検索の機能や、部分試聴の機能を新たに追加した。さらに、音符を自動で分割することで、ユーザの選択肢を減らし、より操作を簡素化した。これにより、よりシンプルな操作と訳詞作業に特化した機能を充実させ、訳詞の作業環境インタフェースとしての改善を実現した。

5.3 システム実装

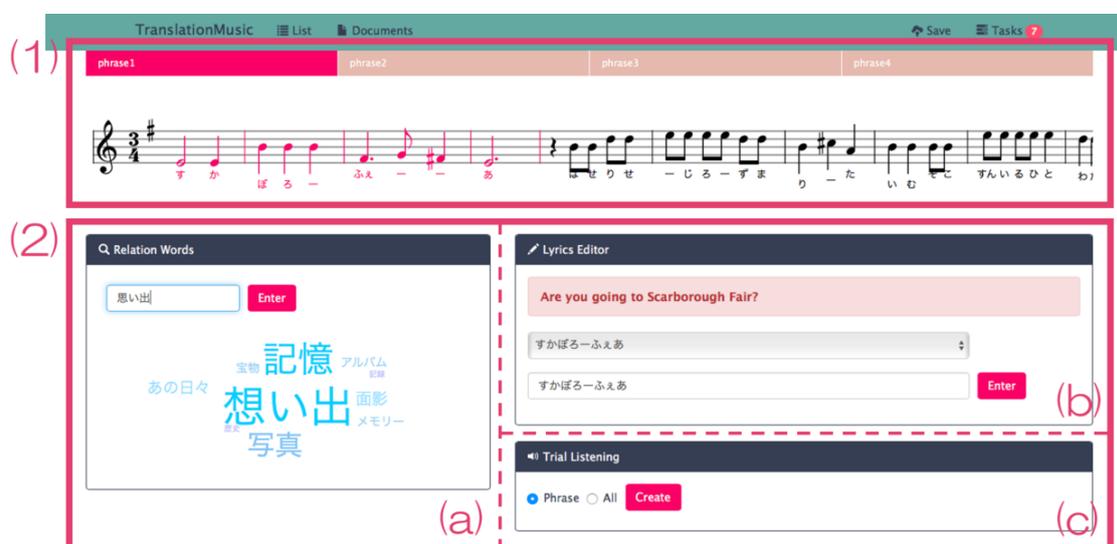


図 5.2: システムインタフェース。

予備実験の結果を取り入れて訳詞システムの改良を行った(図 5.2)。インタフェースは大きく (1) 譜面表示部分と (2) 編集部分の 2 つの要素に分けられる。

予備実装の実験結果を踏まえて、次の機能を追加および改良した。譜面表示部分では、上部にある譜面のフレーズ番号をクリックすることで、編集対象の訳詞フレーズを変更できる機能を追加した。編集部分は、タブ機能を廃止し、(a) 関連語の検索、(b) 訳詞の編集、(c) 訳詞の試聴の 3 つの編集機能を 1 つの表示にまとめた。関連語検索機能は本章の実装で新たに追加された機能である。また、右上のメニューバーに訳詞作業のための、タスクリストを追加した。ユーザがこのメニューをクリックすると、タスクの一覧が記されたモーダルが開く。リストにはグレーのチェックアイコンが表示され、ユーザがタスク

を完遂するとピンクに変わる。最後に、システム全体のテーマとなる色を設定し、インタフェースの表示を整えた。

5.3.1 関連語の検索

図 5.2(2-a) は、入力した単語の関連語を Word Cloud として表示する。関連語はコサイン類似度に基づいて選択され、ユーザがテキストフィールドに単語を入力すると下部に表示される。入力された単語と類似度が高い単語は大きな文字で表示される。本実装では word2vec[Mikolov13] と piapro¹ から作成した歌詞コーパスを用いてベクトル空間を生成した。

5.3.2 訳詞の編集

図 5.2(2-b) は、上からパネル、訳詞候補プルダウン、テキストフィールドで構成されている。パネルには原言語の歌詞が表示され、ユーザが編集した歌詞と比較することができる。訳詞候補プルダウンでは訳詞の候補が 4 つ表示され、選んだ訳詞はテキストフィールドに反映される。ユーザはテキストフィールドで訳詞の編集をひらがなで行うことができる。

入力された訳詞のモーラ数が音符の数より多い場合、メロディの音符は自動的に分割される。また、少ない場合、音符は結合される。音符の分割と結合には 3 章で示したアルゴリズムがランダムに適用される。

5.3.3 試聴

予備実装から試聴機能を拡張し、試聴対象を 1 フレーズまたは曲全体から選択できるように改良した。本章のシステムでは、Sinsy[Oura10] を引き続き使用している。

5.4 手作業と提案システムによる訳詞作業の比較評価

提案システムを用いた訳詞作業と手作業による訳詞作業を被験者に依頼し、それぞれの作業環境について 5 段階で評価してもらった。また、web 検索エンジンを用いることは日常的な手段であるため、どちらの場合においても、被験者は訳詞の作業を行いながら web 検索を行うことを許可した。訳詞作業の制限時間は設けず、被験者は好きなだけ作業を行うことができたようにした。

続いて、提案システムと手動による訳詞作業の実験手順について述べる。提案システムによる訳詞の手順は以下の通りである。

¹<http://piapro.jp>

1. 原曲を聴き，原曲の歌詞を見る．
2. 手作業で訳詞の作業を行う (5分)．
3. 提案システムの使い方を学ぶ．
4. 提案システムの使い方を練習する (5分)．
5. 提案システムを使って，訳詞の作業を行う．
6. アンケートに回答する．

手作業による訳詞の手順は以下の通りである．

1. 原曲を聴き，原曲の歌詞を見る．
2. 提案システムの使い方を学ぶ．
3. 提案システムの使い方を練習する (5分)．
4. 手作業による訳詞作業を行う．
5. アンケートに回答する．

手作業で訳詞を行う被験者には，譜面と歌詞が書かれた紙と譜面だけが書かれた紙の2枚を配布した．

実験では「Scaborough Fair」の楽曲を使用した．また，手作業による訳詞作業の体験やシステムの説明には「Lavendar's Blue」の楽曲を使用した．

提案システムでは，機能ごとに被験者のマウスクリック回数を記録した．また，提案システムと手作業の両方において，作業時間を記録した．

作業後に，定性評価のためのアンケートを実施した．被験者には，以下の4つの項目に対して5段階のリッカート尺度を用いて，提案システムによる訳詞と手作業による訳詞作業の相対評価をしてもらった．

- 日本語と英語の比較のしやすさ
- 日本語の想像しやすさ
- 訳詞作業のしやすさ
- 完成した曲への満足度

5.4.1 結果

今回の比較実験では、楽譜を読む被験者のみを集めた。参加者のうち6名には提案システムを用いて訳詞の作業をするよう指示し、残りの6名には手作業で訳詞をするよう指示した。参加者は女性9名、男性3名（21歳から54歳、平均年齢29歳）であった。また、アンケートでは、被験者に自分の英語能力を5段階のリッカート尺度で自己評価してもらった。提案システムを使用した被験者の平均点は3.5点であり、手作業で訳詞を行った被験者の平均点は3.0点であった。

なお、本実験では、手作業で訳詞を行う作業を行った数名の被験者が訳詞を断念したため、作業時間については考慮しない。

システムログの結果

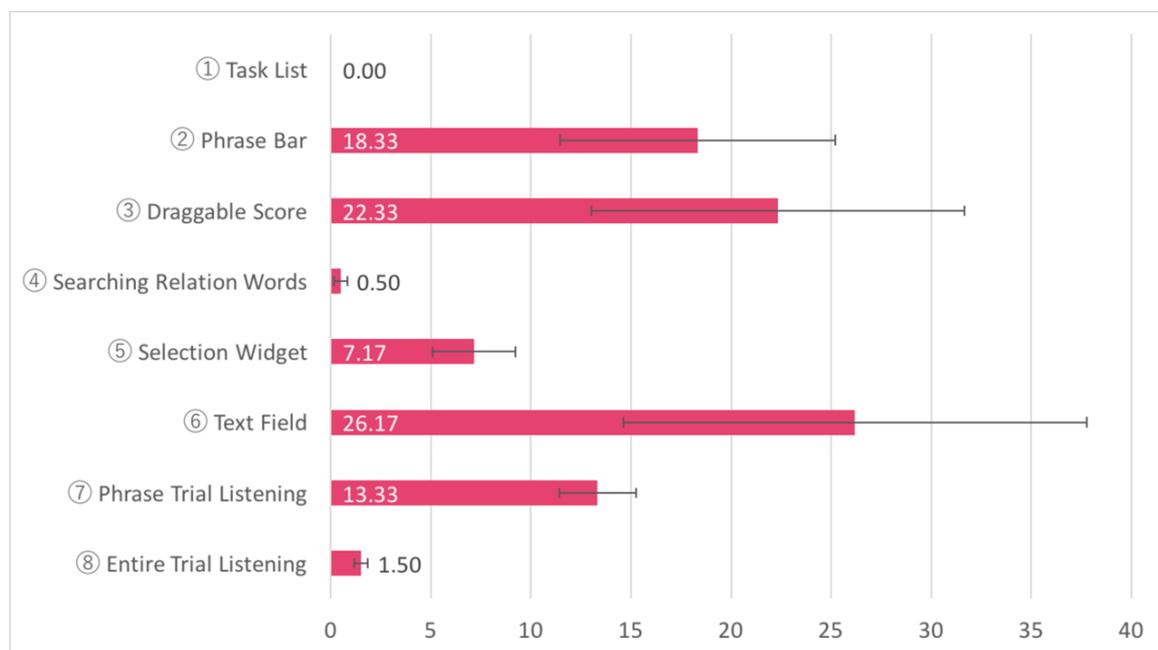


図 5.3: 各機能の操作回数.

提案システム各機能の平均操作回数と標準誤差を図 5.3 に示す。縦軸上の番号は図 5.4 に示す操作可能な部分の番号に対応している。

被験者のほとんどが、タスクリストや関連語の検索を行わなかった。実験開始前にシステムの使い方を説明し、被験者から質問を受けながらシステムの練習を行ったため、被験者はタスクリストを使用する必要がなかったと考えられる。また、本実験では被験者にウェブ上の検索エンジンの使用を許可したため、関連語検索機能が使われなかったと推測できる。提案システムを利用した被験者のウェブ上の検索エンジンでの検索回数の平均は 2.6 回であったのに対し、手動の被験者の平均検索回数は 3.0 回であった。

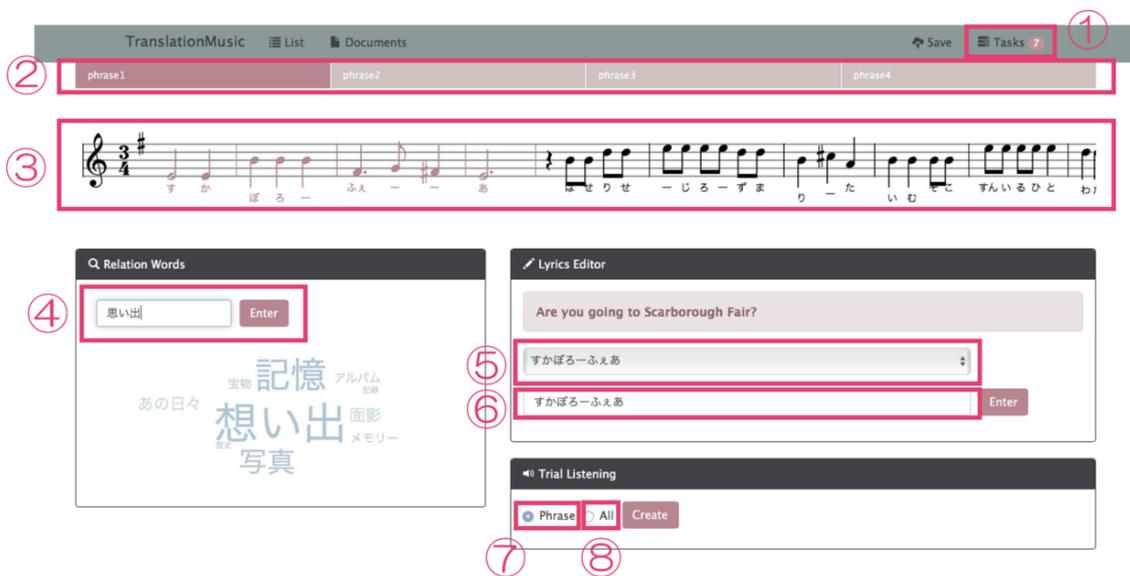


図 5.4: 操作可能部分.

クリック操作によるフレーズ切り替えの平均回数は 18.33 回，ドラッグ操作によるスコア切り替えの平均回数は 22.33 回であった．この結果から，被験者がフレーズを切り替える際にドラッグした回数とほぼ同数，フレーズバーをクリックしており，クリックでフレーズを切り替える機能は有用であることが示唆された．

フレーズ部分試聴機能は平均 13.33 回クリックされ，全曲試聴機能の平均 1.50 回よりも頻度が高かった．全曲視聴機能は最終的な確認をする場面で利用する被験者が多かった．このことから，訳詞のインタフェースには，フレーズごとの試聴が不可欠であることが示唆された．

比較結果

ここでは，提案システムと手作業による訳詞作業のグループそれぞれに同じ設問を設定したアンケートの比較結果について延べる．5 段階のリッカート尺度を用いたアンケートに対する平均値，標準誤差，t 検定の結果を図 5.5 に示す．提案システムはすべての評価項目において高い評価を得た．また，4 項目のうち 3 項目で有意差が認められた．

日本語歌詞と英語歌詞の比較のしやすさという点では，提案システムの標準誤差は，手作業による訳詞作業の標準誤差よりも小さかった．これは，提案システムを用いることで，インタフェースが歌詞を比較するための一定の表示スタイルを提供するため，より効果的に歌詞を比較できたためだと考えられる．これに対して，手作業での被験者は，白紙に直接歌詞を書いたり，英語の歌詞の下に，自分の訳詞を書き込んで歌詞を作成するなど，様々なスタイルを示していた．

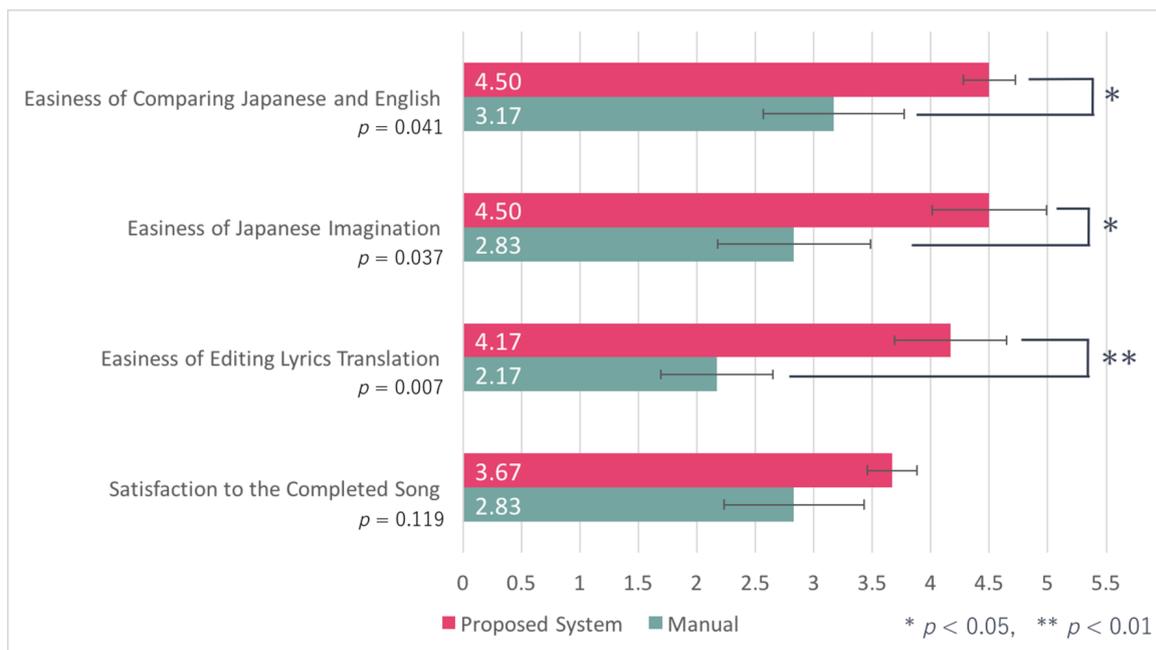


図 5.5: 比較実験のアンケート結果.

日本語での歌詞のイメージのしやすさという点では、「システムが訳詞の候補を提示してくれるので、システムを使いながら日本語で考えるのが楽だった」という意見が複数あった。

また、訳詞の編集のしやすさは $p < 0.01$ となった。今回の実験は、楽譜が読むことが出来る被験者を集めて行った。そのため、手作業による訳詞においても、全く譜面が読めず訳詞の作業が進行できないという事態は発生しなかった。しかしながら、この結果から、楽譜を読むことができるユーザにとっても、手作業で訳詞作業を行うより、提案システムを利用する方が訳詞が容易であることがわかった。

完成した曲に対する満足度は、標準誤差に対して明らかな差が見られた。提案システムを用いて訳詞を行った被験者は、確実に最後まで訳詞の作業を行うことができた。一方、手作業で訳詞を行った被験者は、作業を断念した被験者が複数いた。その結果、作業を断念した被験者は、手作業での訳詞作業に対する満足度の評価を著しく下げる結果となった。そのため、標準誤差に明らかな差が生まれたと考えられる。また、提案システムの評価は、他の評価項目と比較してやや低めであったため、満足度の向上が今後の課題である。

比較実験で得た訳詞のモーラ数について、モーラ誤差率である RMSE を算出した。RMSE の値はシステムの訳詞が 4.52、手作業の訳詞が 4.69 であり、大きな差は見られないがシステムの誤差の方が小さい結果となった。また、アルゴリズムによる訳詞候補の RMSE は 5.27 であったが、ユーザによる編集により、原曲のメロディと訳詞のモーラの差を小さくできることがわかった。

被験者からのコメント

被験者から、提案システムと手作業での訳詞について、次のような意見を得られた。

提案されたシステムについての肯定的なコメントは以下の通りである。

- 編集した曲をリアルタイムで試聴することができて良かった (2名)。
- 訳詞候補を使って歌詞を編集できるのが便利 (4名)。
- 音符を自動で分割、結合して割り当ててくれるのが良い。

提案システムについての否定的なコメントは以下の通りである。

- 音符の分割を自分で変更したかった (3名)。
- 音符の割当のアルゴリズムがランダムのように見え、ルールが無いように見えた。
- YouTube で原曲を聞きながら作業していたので、原曲と編集した曲を聴き比べる機能が欲しかった。

4人の被験者が訳詞候補の機能が便利であると回答した。音符の編集方法については、複数の参加者が「自動編集が便利」とコメントする一方、「手動で編集したい」というコメントもあるなど、意見が分かれた。音符の自動編集では、2つのアルゴリズムのうち1つをランダムに適用するため、どのような音符が割り当てられるか予想ができず、何人かの被験者に対してストレスを与える結果になったと考えられる。

手作業に関する肯定的なコメントは以下の通りである。

- 歌詞の書き換えが簡単だった。
- 英語の歌詞を元に訳詞を編集できた。
- 訳詞の編集箇所がわかりやすかった。
- 音符の分割を自由にできた。
- 漢字を使えるのが良い

手作業についての否定的なコメントは以下の通りである。

- 歌詞を書き換えるのが難しかった。
- 編集した曲の音をイメージするのが難しかった (2名)。
- 歌詞と音符が一致しなかった。

手作業での訳詞作業について、歌詞の書き換えが簡単だったという意見と、難しかったという意見があった。また、手作業では音をイメージすることが難しかったという意見があったが、提案システムには試聴機能が実装されており、この課題は解決できている。また、手作業での訳詞作業では、被験者が歌詞と音符を一致させることが困難であったが、提案システムでは音符の自動編集機能があるため、被験者は割当てを自身で考える必要はない。

完成した曲についての被験者のコメントは以下の通りである。

提案システムで完成した曲

- 訳詞をすることができたが、もっと音符と歌詞の対応を最適化する能力を高くしたいと思った(2名)。
- 歌詞がつまりすぎた。
- 編集した歌詞が訳詞候補に左右される。
- 短時間で良い歌詞が作れた(2名)。
- 英語は分からないけど、システムを使って訳詞できたのが楽しかった。

手作業で完成させた曲に対するコメントは次の通りである。

- 残念な歌詞になった。
- 歌詞ではなく、直訳だった(3名)。
- 訳詞に対して才能が無いと感じた。

提案システムを使用した被験者は、概ね完成した訳詞に対して肯定的な感情を持っていることが伺える。しかしながら、完成した曲の歌詞が、提案システムの訳詞候補に依存しまうとの意見があった。

手作業で訳詞を行ったグループでは、数名の被験者が手作業での訳詞の歌詞は直訳に近いと述べた。一方で、提案システムを使った被験者は自分の歌詞をより音符に合わせたいという回答をしていた。手作業の訳詞の被験者は直訳に悩んでいたのに対し、提案システムを使った被験者は、メロディに合った歌詞にするにはどうすれば良いかを考慮していた。これは、提案システムでは、訳詞が直訳になってしまう問題をすでに解決できている、その先の歌詞をメロディに合わせる段階に進んでいることを示唆している。

専門家からの意見

提案システムについて訳詞の専門家2名に意見を伺った。コメントは以下の通りである。

- 訳詞作業にはとても便利なシステムだと思う。
- 合成された歌は、日本語できちんと歌詞を聴くことができれば確認に役立つ。
- 関連語検索機能はメロディに合う類義語を辞書を見ながら探すことが多いので助かる。
- 複数の歌詞を比較したいので、歌詞をデータとして保存したい。
- 関連語検索機能は関連語だけでなく、類義語も調べられると良い。
- 訳詞によってリズムが変わるので、音符は手入力できると良い。
- 二重母音や無声音など、複数のひらがなを1つの音符に割り当てたい。

専門家たちは提案システムに対しておおむね好意的な意見を挙げてくれた。しかしながら、音符の入力とひらがなの割当については議論が必要である。

5.4.2 考察と議論

比較実験の結果、以下の結論が得られた。

- 楽譜を読むことができるユーザにとっても、提案システムの方が手作業よりも訳詞作業がやりやすかった。
- 関連語検索機能、タスクリスト機能はほとんど使用されなかった。
- 完成した訳詞は、提案システムが提供する訳詞候補に依存していた。
- 提案システムでも歌詞と音符の最適化が困難であった。
- モーラ数が多い仮の訳詞が表示されても、ユーザが編集することによりモーラ数を減らすことができた。

提案システムは、試聴機能、訳詞候補機能、自動音符分割機能を備えていることで、手作業による訳詞の問題を解決することができることが分かった。一方で、いくつかの問題点も挙がった。クリック数の結果から、関連語検索機能とタスクリスト機能は不要であることが分かった。被験者は関連語を検索したい場合、提案システムの機能ではなく、Web 検索エンジンを利用していた。しかしながら、専門家からは関連語検索は有用であるとの回答が得られたため、関連語の表現方法を変更し、検索精度を向上させることで、関連語の利用回数が増えるかどうかを評価する必要がある。

もう一つの問題は、完成した歌詞が訳詞候補に依存することである。提案システムでは、ユーザが仮訳を読み、関連する単語を検索することで、自身の意図や文化的文脈を訳詞に反映させることを期待した。しかしながら、ユーザは提示された訳詞候補に思考を左右されるため、システムが文化的文脈を解釈し、訳詞に反映させることができなければ、これを達成することは難しい。たとえば、この実験で使われたスカボロー・フェアの歌詞には、「パセリ、セージ、ローズマリー、タイム」という4つのハーブの名前が登場する。西洋の文化的文脈では、これらのハーブは中世のお守りと考えられていたが、日本語に訳詞をする際に、この文化的文脈を反映した参加者はいなかった。これは、使用者が仮訳から文化的背景を推測できず、ハーブは植物を意味するだけだと考えていたためである。また、機械翻訳が原言語の文化的背景を無視する問題は、以前から議論されている [Wang13][Hardmeier15][Rozmyslowicz14]。今後は、関連語検索機能など、実装されている機能に加え、LLMを用いた機能を搭載し、ユーザの創造性を刺激する機能を開発していきたい。また、原語の歌詞の文化的背景を訳詞候補に取り入れる方法も検討する必要がある。

現在の実装では、音符の自動編集には2つのアルゴリズムのうち1つをランダムに適用している。この方法だと、どちらかのアルゴリズムがランダムに適用されるという不規則性により、ユーザにストレスを与えてしまう。そのため、音符の自動編集機能のルールを改善する必要がある。また、ユーザが自由に音符を編集するのか、自動編集を適用するのか、さらなる検討が必要である。

5.5 手作業による訳詞とシステムを用いた訳詞の印象評価

比較実験で得た、システムで行った訳詞の結果と手作業による訳詞のそれぞれ6つの結果から、無作為に3つを選択し、合計6つに対して第三者による評価を行った。設問は以下の2種類である。

質問1 次に示す訳が元の歌詞の意味を逸脱しすぎていないか5段階で評価してください。

質問2 原曲の音源と訳詞した音源を最後まで聞いてください。聞いた後、訳詞した音源が原曲の音源に近い雰囲気であるかどうかを5段階で評価してください。

質問1では、原曲の歌詞と被験者の訳詞を文章として並べ、5段階のリッカートスケール(1: 意味を逸脱しすぎている, 5: 意味は逸脱していない)で評価してもらった。質問2では、原曲の音源、被験者が訳詞した音源、の順で曲を聞いてもらい、5段階のリッカートスケール(1: 原曲からは程遠い雰囲気である, 5: 原曲にとっても近い雰囲気である)で評価してもらった。

また、順序効果を軽減するため、質問は回答者ごとにランダムに提示した。

5.5.1 結果

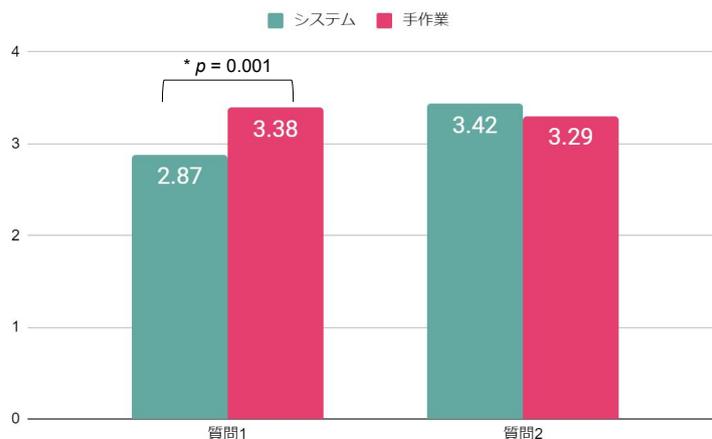


図 5.6: 訳詞の第三者による評価の結果.

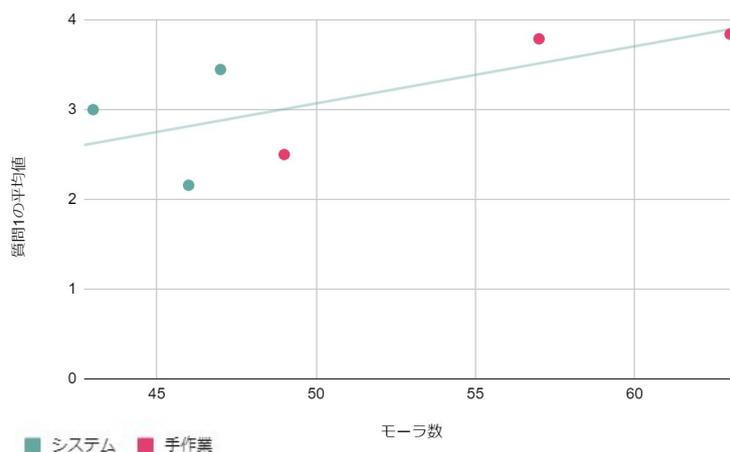


図 5.7: 質問 1 の平均値とモーラ数.

web 上でアンケートを実施し、回答者は女性 37 名、男性 1 名の合計 38 名だった。年代の内訳は、10 代 2 名、20 代 27 名、30 代 8 名、40 代 1 名となった。

質問 1 と質問 2 において、それぞれシステムと手書きによる訳詞ごとに平均値を求めた結果を図 5.6 に示す。質問 1 におけるシステムの訳詞の平均値は 2.87、中央値は 3.00、標準偏差は 1.23 であり、手作業の訳詞の平均値は 3.38、中央値は 4.00、標準偏差は 1.31 であった。質問 2 におけるシステムの訳詞の平均値は 3.42、中央値は 3.00、標準偏差は 1.14 であり、手作業の訳詞の平均値は 3.29、中央値は 3.00、標準偏差は 1.16 であった。また、それぞれの質問に対して t 検定における片側検定を行った。その結果、質問 1 では $P = 0.001$ となり有意差が認められたが、質問 2 では $P = 0.161$ となり有意差は認められな

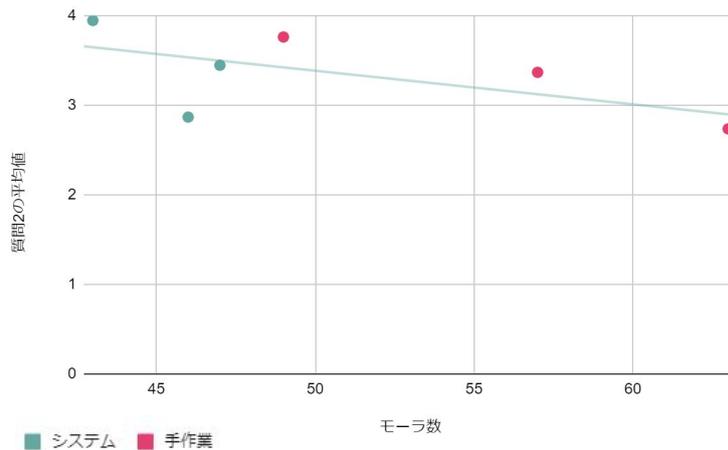


図 5.8: 質問 2 の平均値とモーラ数.

かった。

図 5.7に、質問 1 における被験者による 6 つの訳詞のそれぞれの平均値とモーラ数を示す。モーラ数は 1 曲の全体のモーラ数の合計値を表している。相関係数は 0.70 であり、モーラ数が増加するにつれて、意味を逸脱していない訳詞であると評価される相関関係が確認された。

図 5.8に、質問 2 における被験者による 6 つの訳詞のそれぞれの平均値とモーラ数を示す。相関係数は -0.60 であり、モーラ数が増加するにつれて、原曲からは遠い雰囲気であると評価される相関関係が確認された。

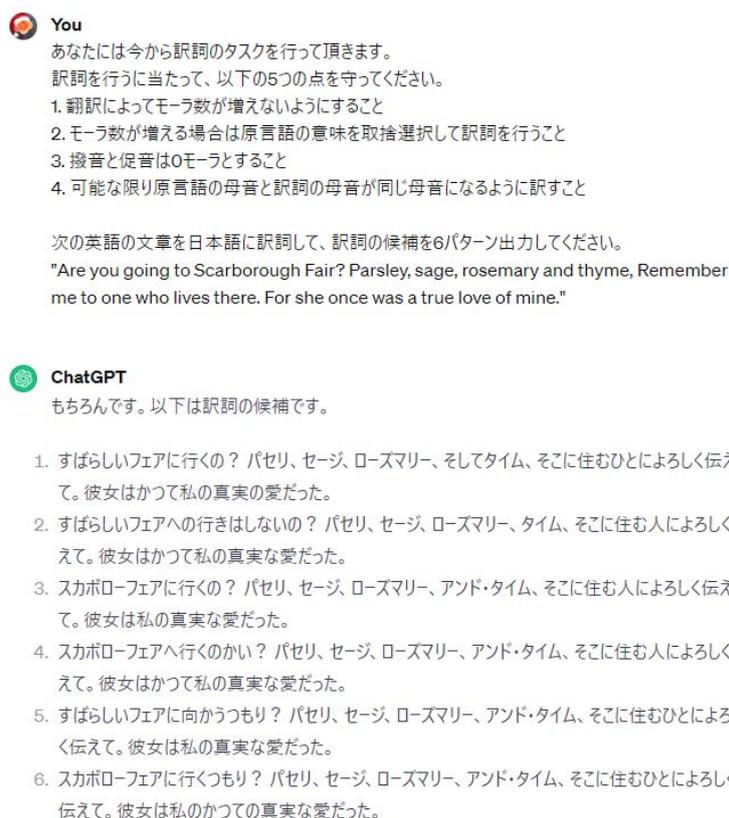
5.5.2 考察と議論

質問 1 において、手作業の訳詞の方が意味が逸脱していないという結果になった。しかしながら、意味が逸脱していないと評価された訳詞はモーラ数が増加する傾向に正の相関があることも分かった。訳詞をする際はメロディの音数は決まっているため、モーラ数の増加は避けなければならない。実験の結果より、質問 2 においてモーラ数と原曲の雰囲気に関して負の相関関係が確認されたように、モーラ数が増加すると原曲の雰囲気は再現できない傾向にある。3.2節にて述べたように、原曲の雰囲気を損なわないようにするためには、歌詞の意味を変更または省略してでもモーラ数を減らす必要があることが分かった。以上より、訳詞においては歌詞の意味と曲の雰囲気はトレードオフの関係にあることが示唆された。

歌詞の意味を取るか、曲の雰囲気を取るかは訳詞者の感性に委ねるべき部分である。そのため、システムにはどちらをより重要視するかについてのパラメータを組み込むべきだろう。また、現状のシステムでは訳詞アルゴリズムから得た訳詞候補をユーザが編集することで、モーラ数を減らすことが出来る

が、より原曲の雰囲気近づけるためにはモーラ数の差分を0に近づける必要がある。モーラ数を減らすためには、主語の省略、フレーズ間での意味の入れ替えなども有効な手段であるため、今後の実装に取り入れていきたい。

5.6 モーラ誤差率と文章類似度による訳詞の定量評価



You
あなたには今から訳詞のタスクを行って頂きます。
訳詞を行うに当たって、以下の5つの点を守ってください。
1. 翻訳によってモーラ数が増えないようにすること
2. モーラ数が増える場合は原言語の意味を取捨選択して訳詞を行うこと
3. 撥音と促音は0モーラとすること
4. 可能な限り原言語の母音と訳詞の母音が同じ母音になるように訳すこと

次の英語の文章を日本語に訳詞して、訳詞の候補を6パターン出力してください。
"Are you going to Scarborough Fair? Parsley, sage, rosemary and thyme, Remember me to one who lives there. For she once was a true love of mine."

ChatGPT
もちろんです。以下は訳詞の候補です。

1. 素晴らしいフェアに行くの？ パセリ、セージ、ローズマリー、そしてタイム、そこに住むひとによりしく伝えて。彼女はかつて私の真実の愛だった。
2. 素晴らしいフェアへの行きはしないの？ パセリ、セージ、ローズマリー、タイム、そこに住む人によりしく伝えて。彼女はかつて私の真実な愛だった。
3. スカボローフェアに行くの？ パセリ、セージ、ローズマリー、アンド・タイム、そこに住む人によりしく伝えて。彼女は私の真実な愛だった。
4. スカボローフェアへ行くのかい？ パセリ、セージ、ローズマリー、アンド・タイム、そこに住む人によりしく伝えて。彼女はかつて私の真実な愛だった。
5. 素晴らしいフェアに向かうつもり？ パセリ、セージ、ローズマリー、アンド・タイム、そこに住むひとによりしく伝えて。彼女は私の真実な愛だった。
6. スカボローフェアに行くつもり？ パセリ、セージ、ローズマリー、アンド・タイム、そこに住むひとによりしく伝えて。彼女は私のかつての真実な愛だった。

図 5.9: ChatGPT に与えた入力と出力。

比較実験で得られた、システムで行った訳詞、手作業による訳詞のそれぞれ6つの結果に加えて、LLMを用いたサービスの訳詞の精度を測るため ChatGPT 3.5 を用いて作成した訳詞に対して、モーラ誤差率と翻訳スコアを用いた評価を行った。図 5.9に ChatGPT に与えたプロンプトの入力とその結果を示す。比較実験で得た訳詞の個数と同等になるように6種類の訳詞を出力するように指示した。

原言語と訳詞のモーラ誤差率は、RMSE を用いて算出した。また、翻訳スコアとして、多言語埋め込みモデルである Languageagnostic BERT Sentence Embedding(LaBSE)[Feng22] と Multilingual Embeddings from bidirectional Encoder representations(ME5)[Liang22] を用いて、原言語の歌詞と訳詞の cos 類似度を文章類似度としてフレーズごとに算出した。なお、翻訳のタスクにおいては一般的に BLEU スコアが指標として用いられるが、Scarborough Fair の楽曲に日本語訳詞が存在しないため参照となる訳を決

定できない点と、訳詞においては意識を行うのが一般的なため N-gram の一致度を考慮する BLEU は適切な指標ではないと判断し、今回は評価指標から除外する。

5.6.1 結果

表 5.1: 各訳詞の RMSE および文章類似度.

	RMSE	LaBSE	ME5
システム	4.52	0.59	0.83
手作業	4.69	0.52	0.81
ChatGPT	6.94	0.72	0.86

システムで行った訳詞、手作業で行った訳詞、ChatGPT による訳詞に対して RMSE, LaBSE, ME5 を算出した結果を表 5.1 に示す。LaBSE と ME5 は、フレーズごとに算出した値に対して平均値を取った値である。システムの訳詞と手作業の訳詞を比較すると、システムの訳詞の方がモーラ誤差率が低く、LaBSE と ME5 による文章類似度が高いという結果になった。これは、システムの訳詞の方が歌唱しやすく、現言語の意味を日本語の歌詞に反映できていることを示唆している。また、ChatGPT による訳詞は LaBSE と ME5 による文章類似度が最も高いが、モーラ誤差率も高い。この誤差率の内訳として、ChatGPT による訳詞のすべてのフレーズにおいて、モーラ数が増加していた。

5.6.2 考察と議論

表 5.2: 定性評価で用いた評価対象の RMSE および文章類似度.

	RMSE	LaBSE	ME5
システム	3.57	0.56	0.82
手作業	5.87	0.57	0.82

実験の結果より、システムを使った訳詞は手作業の訳詞と比較して、モーラ誤差率が低くなることが分かった。また、LaBSE と ME5 を用いた文章類似度の評価指標の両方において、システムの訳詞の方が原言語との文章類似度が高かった。

一方で、5.5 節の実験ではシステムの訳詞の方が意味が逸脱していないという結果が出ていた。そこで、5.5 節の実験で評価対象だったシステムの訳詞 3 つと、手作業の訳詞 3 つに対して文章類似度を算出した結果を表 5.2 に示す。ME5 は同等の値であるが、LaBSE については手作業の訳詞の方が文章類似度が高くなったため、今回の実験で用いた指標においても 5.5 節の実験と同じ結果になったと見なせる。この結果より、システムを利用して訳詞作業を行うと、モーラ数の増減を抑えながら、原言語の意味を反映した訳詞を行えることが示唆された。

ChatGPTを使った訳詞については、文章類似度は高かったが、訳詞のモーラ数も増加していた。5.5.2節の議論で述べたように、訳詞においてはモーラ数が増えると意味が逸脱しづらいが、原曲の雰囲気とは遠くなる、というトレードオフの関係にある。そのため、ChatGPTによる訳詞は原曲の雰囲気を再現できないと考えられる。このことから、システムの訳詞とChatGPTの訳詞を比較したときに、システムの訳詞の方が原曲の雰囲気を保った訳詞であると判断できる。

結論として、本論文で提案した訳詞システムは、手書きやChatGPTを用いた訳詞と比較して、原曲の雰囲気を保ちつつ意味を反映した訳詞を作成できることが示唆された。

5.7 まとめ

本章では、予備実験の結果からシステムを改善し、5.3節で改良したインターフェースについて述べた。システムと手作業での訳詞作業の比較実験(5.4節)では、12名の被験者による評価を行った。その結果、訳詞システムは、日本語と英語の比較のしやすさ、日本語の想像しやすさ、訳詞作業のしやすさの点で、手作業よりも有意に良い評価を得られた。一方で、完成した曲への満足度は改善が必要になることが分かった。また、比較実験で得た訳詞が原曲の意味や雰囲気を表現できているかを38名に評価してもらった(5.5節)。その結果、訳詞においては歌詞の意味と曲の雰囲気はトレードオフの関係にあることが示唆された。5.6節では、RMSEによるモーラ誤差率、LaBSEとME5による文章類似度を用いた評価を行った。システムを用いて作成した訳詞は、モーラ誤差率が低く、文章類似度が高いという結果が得られた。これより、本論文で提案した訳詞システムは、原曲の雰囲気を保ちつつ、現言語の意味を反映した訳詞を行えるインターフェースであることが示唆された。

今後は、歌詞の文化的文脈を反映できるような機能の追加や、メロディと歌詞のマッチングにユーザの意図を反映できるよう改善を行いたい。

第6章 iMake: 役のモチーフに基づいたアイメイクデザインを生成するシステムの試作と予備実験

6.1 はじめに

本章では、役のモチーフに基づいたアイメイクデザインを生成する上で、提案インターフェースの有効性や転写シールによる化粧の実現可能性を検証するための試作について説明する。6.2章にて舞台演劇におけるアイメイクの課題点について述べる。6.3章で役のモチーフに基づいたアイメイクデザイン生成を補助するシステムの試作について述べ、6.4章にて試作したシステムの予備実験と評価について述べる。

6.2 舞台演劇におけるアイメイクの課題

本システムは舞台演劇においてアンサンブルと呼ばれる、コーラスやダンスなどの複数の役をこなす役割を持った役者向けのシステムである。

目の魅力は顔の魅力と最も強く相関しており [Saegusa15]、一般的には目が大きい方が好まれる傾向がある [Morikawa13][Baudouin04]。目を大きく見せるための化粧には、アイブロー、アイシャドウ、アイライン、マスカラなどがあり [Morikawa15][Muto19]、これらを用いた化粧を一般的にアイメイクと呼ぶ。今回は、化粧の中でもアイメイクに注目する。演劇におけるアイメイクは、客席からはっきりと顔が見えるように、アイシャドウは濃く塗り、アイラインは目の大きさを大幅に越えて引くため、通常アイメイクとは異なる。また、アンサンブルは複数人で同じ役を演じる場合、全員同じ化粧のデザインを施すことがあるが、化粧の技量は個人によって差があるため、同一のデザインでも観客からの見え方が異なる場合がある (課題 1)。加えて、舞台上では灰色、黄色、青、紫など、通常のメイクでは使用しない色を用いることも多い (課題 2)。

また、自分に似合うメイクではなく、役になりきるための化粧のデザインが必要である。そのため、演じる役に応じて、魚の場合は鱗のようなデザイン、蜘蛛の場合は蜘蛛の巣のようなデザインを化粧に取り入れることがある。しかしながら、このようにモチーフの形状や色などの特徴を捉えた化粧のデザインの生成は個人によって得手不得手があると考えられる (課題 3)。

ここで課題をまとめる。

課題1 メイクの技量によって仕上がりに個人差がある

課題2 所持している化粧品によって使用できる色が制限される

課題3 モチーフの色彩や形状の表現を反映した独創性の高いデザインを考案する技量に個人差がある

以上より、舞台化粧は特殊な上、アイメイクのデザインにモチーフの特徴の反映が難しく、個人の技量で仕上がりに差が出るという課題がある。

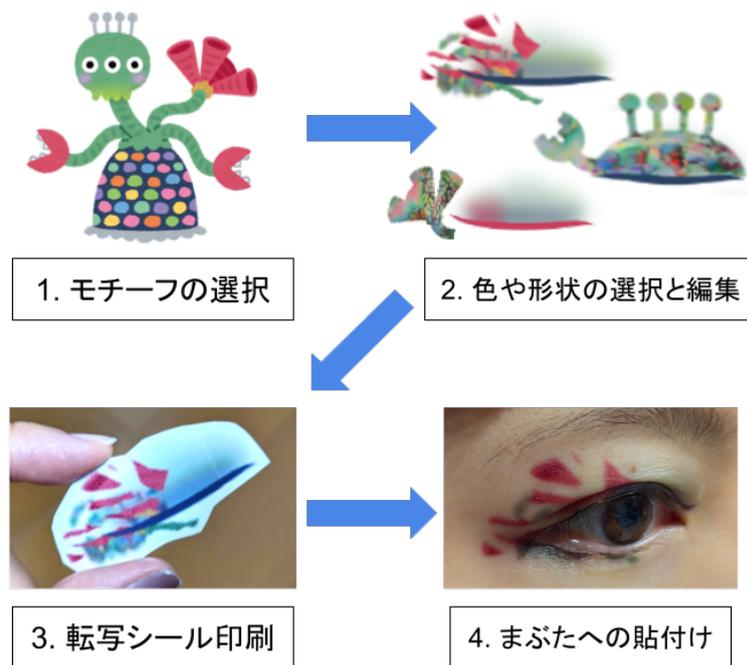


図 6.1: iMake を用いて転写シールで化粧をした例。

本論文では、モチーフとなる画像からアイメイクのデザインを生成し、転写シールに印刷することで、瞼に貼って実際に利用することができるシステム“iMake”を提案する。iMake で作成したデザインを転写シールに印刷し、実際にまぶたに貼り付けるまでの流れを図 6.1 に示す。舞台上では人間以外の生き物や、空想上の生物、概念などの抽象的な役を演じることも多々ある。そのため、存在しない生物なども表現できるイラスト画像から特徴を抽出してデザイン生成を可能にした。

本システムを利用することで、各課題に対して、以下のように解決を目指す。アプリケーション上で作成したデザインをそのまま転写シールに印刷することで、鏡像の影響や線の不揃いさを解消し、課題1を解決する。転写シールを用いることで、従来のメイク方法と比較して技量に関係なく、簡単にモチーフの特徴を取り入れたアイメイクを施すことが可能になり、仕上がりに対しても一定の水準を担保でき

ることが期待される。

また、シアン、マゼンタ、イエロー、黒 (CMYK) のカラーインクジェットプリンタを利用することで、ユーザが所持している化粧品の色に制限されず、自由に色を選択できるようになる。これにより、アイメイクの色彩の自由度を高め、課題 2 を解決する。

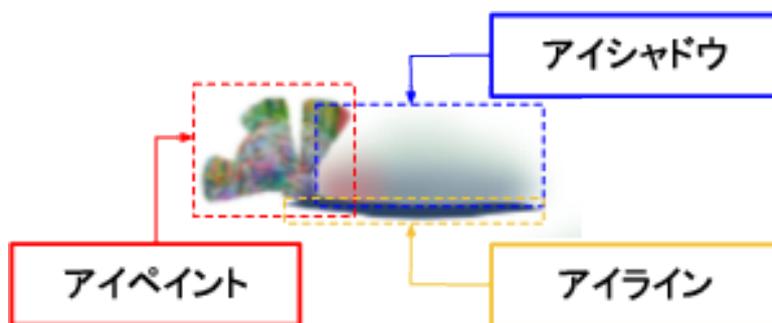


図 6.2: モチーフの特徴を取り入れたアイメイクデザインの 3 つの要素.

さらに本システムでは、モチーフの特徴を取り入れたアイメイクデザインを

- アイシャドウ
- アイライン
- アイペイント

の 3 つの要素に分類した (図 6.2). アイシャドウ, アイラインはアイメイクの用語として一般に使われる単語である。これに加えて, 本論文では図 6.2 に示すような, モチーフの特徴となる形状や模様を表すデザイン部位を「アイペイント」と定義した。ユーザは 3 つの要素をそれぞれ個別に選択・編集するだけで, アイメイクのデザイン作成が可能になる。アプリケーションがモチーフに応じて, アイシャドウ, アイライン, アイペイントそれぞれの要素の色彩や形状を提案することで, デザインの知識がないユーザのアイメイクデザイン作成を補助し, 課題 3 を解決する。本システムでは, 転写シールを使う制限上, アイメイクにおけるアイブロウ, マスカラの過程は実装の範囲外とする。

以上より, ユーザが選択したモチーフから, アイメイクのデザインに特徴を反映し, 転写シールを使うことで個人の技量による仕上がりの差をなくした舞台メイクができるようになることを目指す。

本章では, プロトタイプとして実装したシステムについて述べ, その評価を行う。

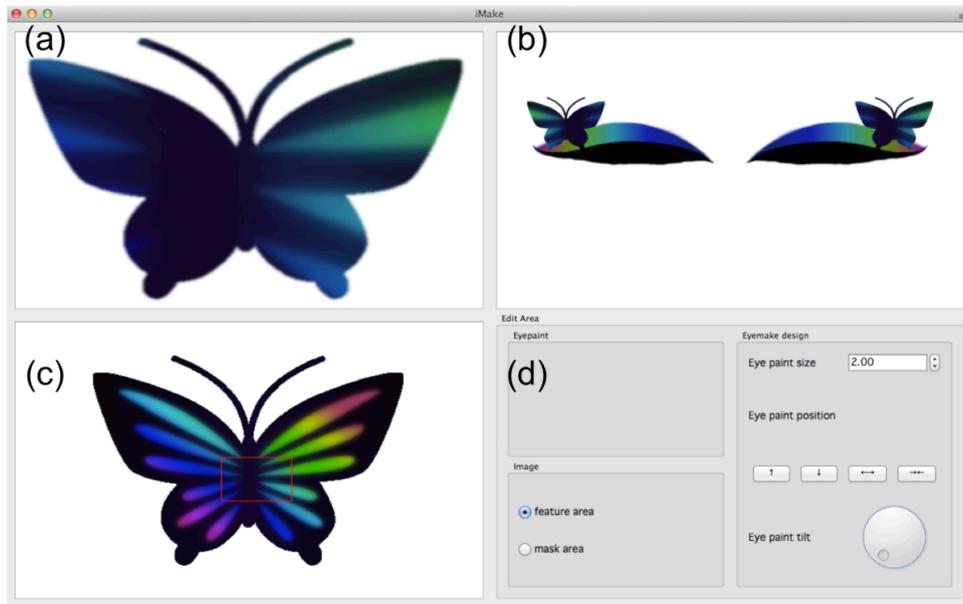


図 6.3: 試作システムのインターフェース

6.3 試作システムのインターフェース

iMake はユーザが用意した画像の全体からアイシャドウのグラデーションを生成する。一方、アイペイントはユーザが矩形で選択した「特徴領域」と「マスク領域」から生成される。矩形選択は最も単純な選択方式であり、難しい操作を必要としないため、これを採用した。

iMake は図 6.3に示すような、4つのパネルにより構成されている。左上のパネルは、アイペイントのデザインを表している (図 6.3-(a))。ユーザはアイペイントの詳細なデザインをこのパネルで見ることができる。右上のパネルは、生成されたアイメイクのデザインを表示している (図 6.3-(b))。ここで生成されたデザインは JPEG ファイルとして保存することが可能である。ユーザはこのファイルを転写シートに印刷することで、自身の顔に貼り付ける。左下のパネルはユーザが選択した画像を示している (図 6.3-(c))。ユーザはこのパネルで特徴領域及びマスク領域を選択する。矩形の種類はラジオボタンによって行われる。これらのラジオボタン、及びアイペイントを編集するためのスピンボックス、ボタン、ダイヤルは右下のパネルに設置した (図 6.3-(d))。スピンボックスはサイズを、ボタンはアイペイントの位置を、ダイヤルは角度をそれぞれ編集することができる。なお、本システムは、OpenCV、OpenGL、C++、及び Qt Creator により PC アプリケーションとして開発した。

続いて、アイシャドウ、アイライン、アイペイントの3つの要素の生成について述べる。

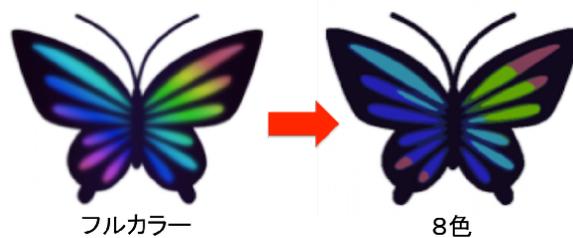


図 6.4: 画像の減色

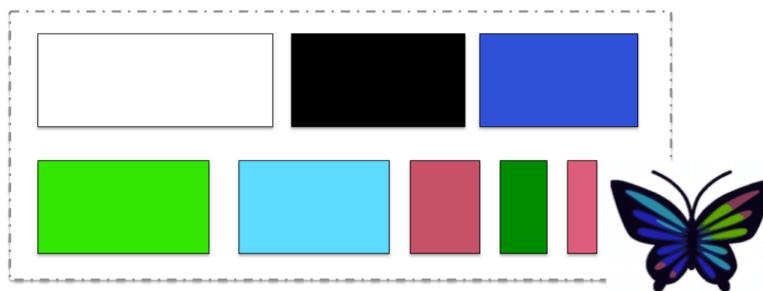


図 6.5: 色の抽出

6.3.1 アイシャドウ

アイシャドウを作成するため、まず初めに、ユーザが提供した画像から代表的な色を 8 色選択する。代表色を 8 色とした理由は、アイメイクを行う際に、通常は 4 - 5 色程度のアイシャドウパウダーを瞼に置き、これをぼかしてグラデーションを作る事から、置いた色とそれぞれの中間の色の数である 7 - 9 色があれば、瞼を覆うのに十分なグラデーションが作れると考えたためである。そこで、図 6.4 に示すように、画像を K-means クラスタリングにより 8 色に減色し、それにより得た色を代表色とした。

通常のメイク方法でアイシャドウを作る場合、テーマとなる色を目の内側に使うことが多い。本システムにおいてもより画素数の多い代表色を画像を表現する色と考えて、より内側に配置することを考えた。しかしながら、画素数順に機械的に並べてしまうと、画像によっては同系色が離散的に配置され縞模様状になり、不自然になる場合があることがわかった。これを避けるため、減色により取得した背景色を含む 8 代表色 (図 6.5) を、まず最初に色のグループ化を行った。このグループは、代表色の RGB 値を HSV 値に変換することで、色相、明度、彩度を判定し、同系色グループに振り分けている。グループは H の値を 20 分割した色彩に加えて、黒、白、灰色の 23 グループを用意した。このグループ化において、白、黒および灰色に振り分けられた代表色については、アイシャドウの色として不適切と考え、以下の処理から除外することにした。

8 代表色を同系色にグループ化した後に (図 6.6)、画素数の最も多い同系色グループを内側に配置する。

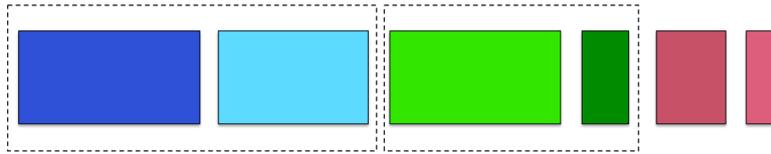


図 6.6: グループに振り分けられた代表色

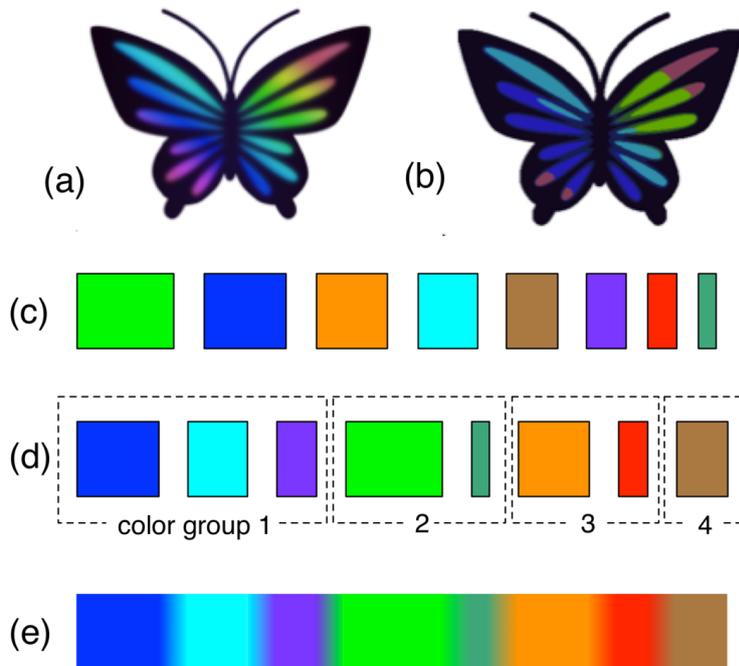


図 6.7: 完成したグラデーション

同じ同系色グループ内に 2 種類以上の代表色が割り当てられた場合には、画素の多い代表色を、グループに割り当てられた領域の内側に配置する。続く代表色は、H の値が近い色を順次配置していく。この時、次に画素数が多い色がある方向に値を進める。これにより、8 色の代表色に対して、色の変化を自然に保ちつつ、色数の多い代表色が内側に配置される傾向にある配置順を実現した。

8 色の代表色の配置順が確定したところで、次にそれぞれの色の表示幅とグラデーションを施す幅を決定する。前述と同じく、画素数の多い代表色は画像を表現する色と考えて、幅の広い領域を割り当てた。この時、隣接する代表色の間に、2 色を接続した際に滑らかに見える最小の幅を設けて、OpenGL のグラデーション機能を使用した (図 6.7)。

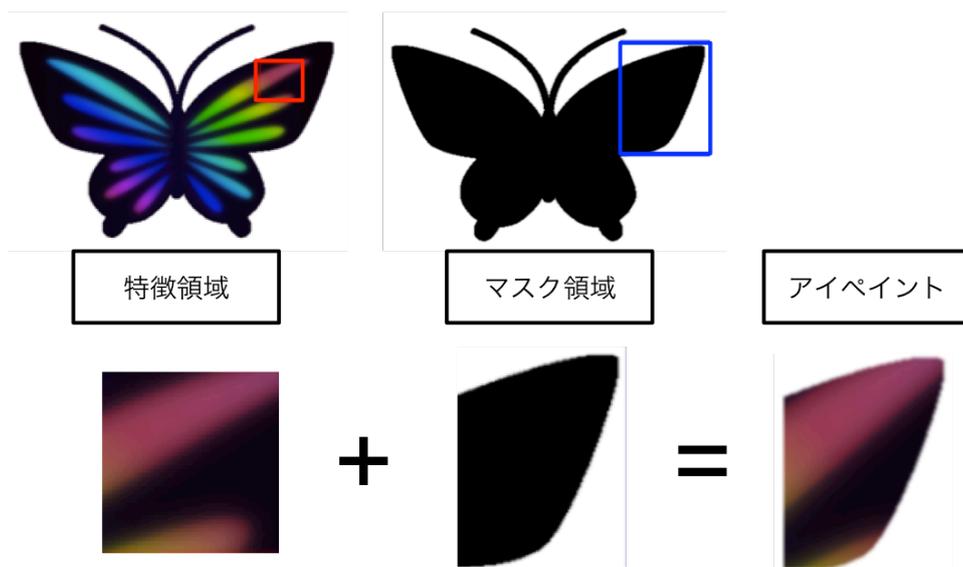


図 6.8: アイペイントの生成

6.3.2 アイライン

iMake では、デザイン中にアイラインを含むことによって、転写シールにアイラインを印刷することが可能である。試作における実装では、アイラインのデザインは一般的な形と黒色の一種類である。

6.3.3 アイペイント

アイペイントを生成する時に、ユーザが矩形として選択した特徴領域とマスク領域を使用する (図 6.8)。システムは、OpenCV のライブラリの機能を利用して、ユーザが選択したもとなる画像 (オリジナル画像) から輪郭を抽出し、閉領域を塗りつぶした画像 (マスク画像) を作成する。

アイペイントは特徴領域の画像をマスク領域の画像サイズへと変形させ、マスクパターンを適用し切り抜くことによって生成される。さらに、本システムではアイペイントのサイズ、位置、角度をユーザが編集することができる。

6.4 試作システムによる予備実験と結果

iMake の有用性を確認するために、試作システムの評価実験を行った。実験に参加した被験者は、22 歳から 27 歳の女性 8 人である。被験者には本システムを使用しデザインを生成した後、それを印刷して実際に顔に貼ってもらった。生成されたデザインの一例を図 6.9 に示す。デザインを生成完了するまでの平均時間は 2 分 29 秒だった。

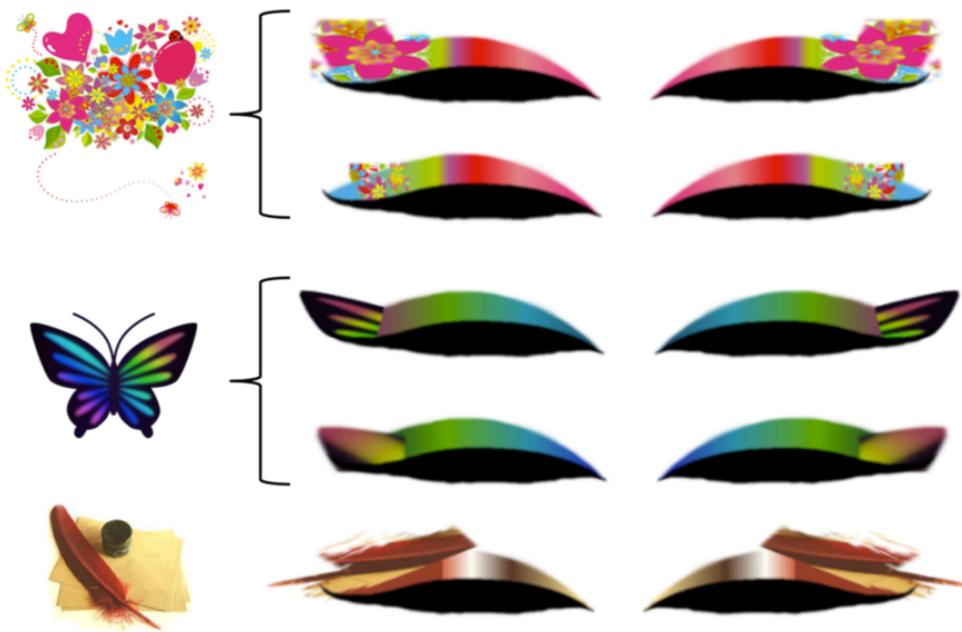


図 6.9: 被験者によって生成されたデザイン

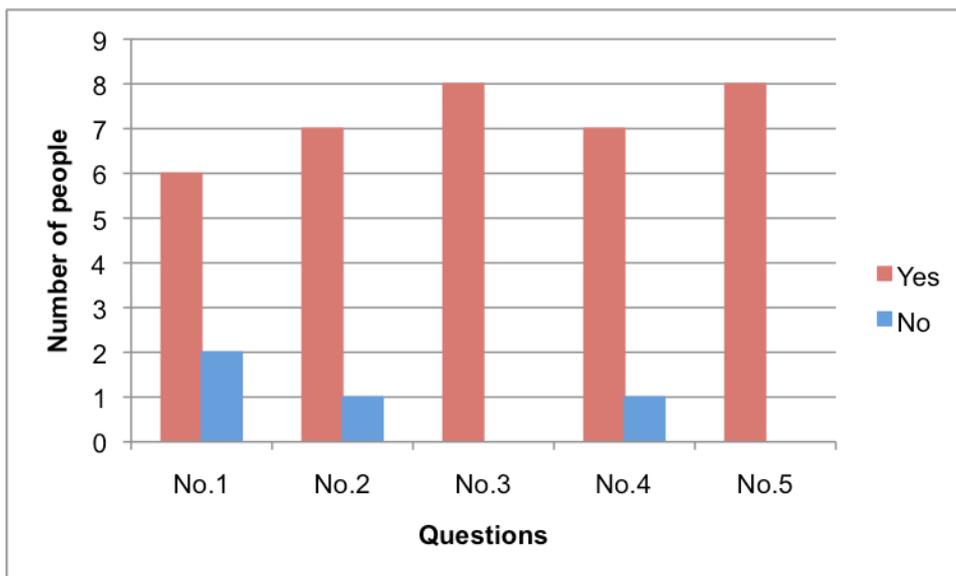


図 6.10: 評価実験の結果 (質問 1-5)

続いて、被験者にはアンケートに答えてもらった。1-2の質問は、被験者のメイクに関する背景を問うものである。残りの3つの質問は、本システムへの評価に関するものである。アンケートの内容を以下に、結果を図 6.10に示す。

1. 普段からメイクを実施しているか？ また、メイクをしている人は週何回の頻度でメイクをするか？
2. モチーフの特徴を反映したアイメイクに興味があるか？

3. 生成されたデザインと元になった画像に関連性は見受けられるか？

4. 転写シールを臉に貼ることは簡単だったか？

5. 生成されたアイメイクのデザインには満足しているか？

8人の内、6人が普段からメイクをしていると解答し、その中の平均は5.2回/週だった(設問1)。また、8人中7人がモチーフの特徴を反映したアイメイクに興味があると解答した(設問2)。全ての被験者が生成されたデザインと元の画像に類似性があるように見受けられたと答えたことから、本システムのデザインアルゴリズム表現が成功していることが示された(設問3)。8人中7人の被験者が、転写シールの適用は簡単だったと答えた(設問4)。難しかったと答えた被験者は、シールを臉のどの位置に置くのか決めるのが難しかったと答えた。この答えから、転写シールを臉に貼るには、個人の技量によって、つけまつ毛のように多少の練習が必要になる可能性があると考えられる。また、全ての被験者が生成されたデザインに満足していると答えた(設問5)。

以下に被験者からの意見を示す。

- システムのインタフェースがとてもシンプルなので、使うのが簡単だった。
- 転写シールを臉に貼った時の感触が、想像していたよりも快適だった。
- アイペイントの編集機能がもっと充実していると良い。ユーザが自分で描いたり消したりするような機能が欲しかった。

6.4.1 考察

予備実験の結果から、普段からメイクをしていないユーザやモチーフの特徴を反映したアイメイクに興味の無いユーザでも、生成されたデザインの満足度が高いことが示唆された。また、試作システムはモチーフ画像の特徴を、アイメイクのデザインに反映できていると考えられる。一方で、被験者によっては、転写シールを臉に貼る行為が難しいと感じているため、通常の化粧と同じように練習が必要であることが分かった。また、より良いデザインを作成するためには、ユーザが自分でアイペイントの形を編集できるような機能が必要であると考えられる。

6.5 まとめ

本章では6.2章で舞台演劇におけるアイメイクの課題を挙げ、6.3章でiMakeの試作システムのインタフェースの内容について述べた。6.4章では、試作システムを用いた予備実験とその結果および考察につ

いて述べた。試作した提案インタフェースの使用により、アイメイクのデザインにモチーフの特徴を反映させることが可能であり、生成されたアイメイクのデザインの満足度が高いことが示された。また、転写シールを用いたアイメイクの貼り付け作業が比較的容易であることが示唆されている。一方で、アイペイントの形状を編集する機能の追加が必要であることが明らかになった。さらに、アイシャドウやアイラインの種類が限られているため、ユーザの選択肢が制限されており、編集機能としては不十分であった。したがって、今後の改良においては、選択肢の拡充を検討する。

第7章 iMake: 役のモチーフに基づいたアイメイクデザインを生成するシステムの実装と評価

7.1 はじめに

本章では、第6章で得た結果をもとに改良したシステムについて述べる。7.2節で改良したインタフェースとシステムの内容について述べ、7.3節では、アイシャドウ、アイライン、アイペイントのデザイン要素に対して、システムがモチーフから提案した内容について評価実験を行った内容と結果について述べる。7.4節にて、従来の化粧方法とシステムを利用した化粧方法による、モチーフの特徴を表現したアイメイクデザインの化粧の比較実験の内容とその結果を述べる。また、比較実験で得た被験者の化粧画像を7.5節の実験にて、第三者による印象評価を行った内容とその結果について述べる。

7.2 システム概要

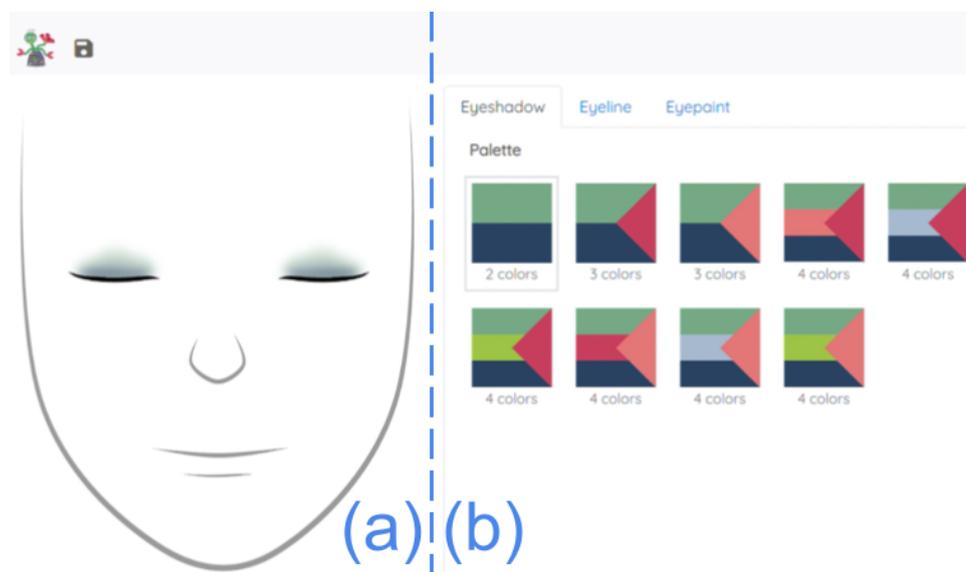


図 7.1: システムインタフェース

本実装で用いるモチーフ画像は、中心部にキャラクターやアイコンなどが描かれた背景が存在しないpng形式の画像とする。ユーザがモチーフ画像を入力すると、デザイン作成用のインタフェースとして

図 7.1が表示される。インタフェースの機能は大きく 2 つに分けられる。

インタフェースの左側 (図 7.1-(a)) は、ユーザが作成しているアイメイクのデザインの全体像を表示するキャンバスである。キャンバス上には顔のガイドラインがあらかじめ表示されており、それに加えてデフォルトで選択されたアイシャドウとアイラインが表示されている。アイシャドウとアイラインは固定の位置・大きさで表示され、アイペイントはキャンバス上で位置や大きさ、形状や角度などを編集できる。

インタフェースの右側 (図 7.1-(b)) には、アイシャドウ、アイライン、アイペイントの 3 つのタブが表示され、タブを切り替えることで、それぞれの要素を編集することができる。試作システムの結果を踏まえ、アイシャドウとアイラインに関しては複数のデザイン候補から色や形状を選択可能とした。さらに、アイペイントでは位置や回転、大きさの編集をドラッグで操作可能にしたことに加えて、消しゴム機能を追加した。形状を編集する機能を消しゴムのみ限定することで、色を選択して書き加えるなどの複雑な操作を排除した。このようにして、本実装では役のモチーフに基づいたアイメイクのデザインのインタフェースを、モチーフの特徴を反映しつつ、よりユーザが様々な表現をしやすいように改良した。

7.2.1 アイシャドウ

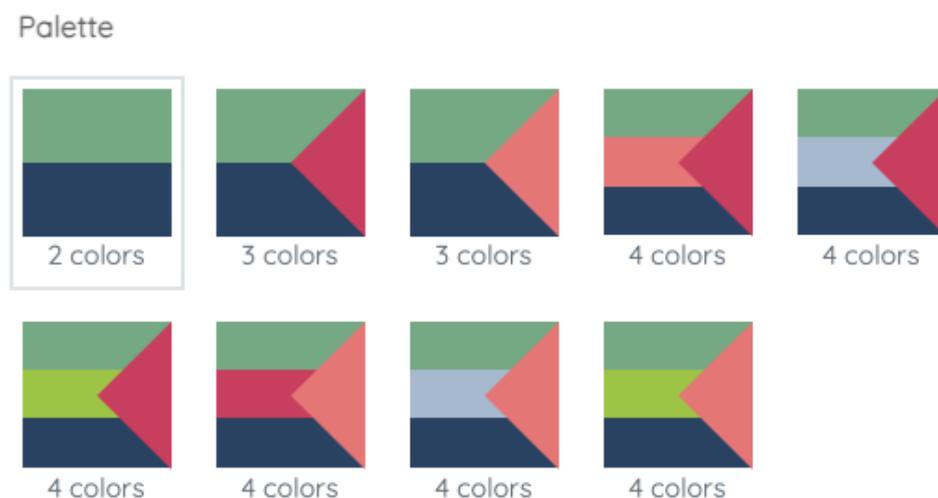


図 7.2: アイシャドウのタブの内容

アイシャドウのタブには、ユーザが入力したモチーフ画像から自動で作成した 2 色、3 色、4 色から構成されるアイシャドウの色の組み合わせ候補が複数表示される (図 7.2)。それぞれの候補は、正方形内で簡易的にアイシャドウのグラデーションの形状を示している。本論文では、この正方形で表された色の組み合わせ候補をパレットと呼ぶ。ユーザが候補の中からパレットを 1 つ選択すると、図 7.1-(a) の顔のグラデーション部分の色と形状が変化する。

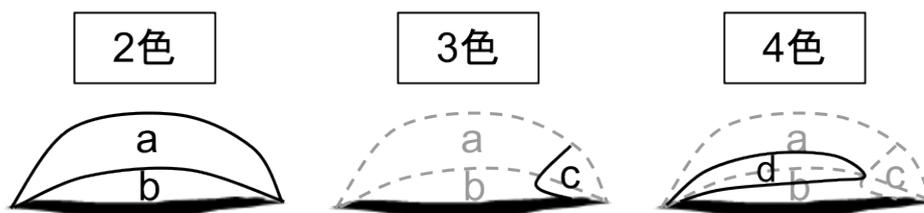


図 7.3: パレットの色数とグラデーション形状

システムが生成したパレットの色数によって、グラデーションの形状は異なる。アイシャドウのグラデーションは、瞼の縁から外側に向かって濃い色から淡い色になるよう色の配置を行う。図 7.3に色の数に対応するグラデーションの形状を示す。現在の実装では、アイシャドウは上瞼のみをターゲットとした。パレットに含まれる色数が2色の場合、アイシャドウのグラデーションは上下2色のグラデーションとなる。このとき、上瞼の眼球と骨の間にあるくぼみ部分の領域を a、瞼の縁に近い領域を b とする。パレットに含まれる色数が3色の場合、領域 a、領域 b に加えて目尻部分に色を追加した3色のグラデーションとなる。このとき、目尻付近の領域を c とする。パレットに含まれる色数が4色の場合、領域 a、領域 b、領域 c に加えて、領域 a と領域 b の間に色を追加した4色のグラデーションとなる。このとき、領域 a と領域 b に挟まれた領域を d とする。今回の実装では、アイラインの色より薄くなるようアルファ値を設定した a, b, c, d の形状に合わせたテンプレート画像を事前に用意し、指定した色と合成することで、グラデーションを作成する。

アイシャドウの色は目の縁側が濃く、縁から離れるに従って色が薄くなるというルールが存在する。これは、通常のメイクにおいても、舞台メイクにおいても変わらない法則である。そのため、アイシャドウの形状および配色アルゴリズムは、通常のアイシャドウの手法を参考に作成した¹²³⁴。

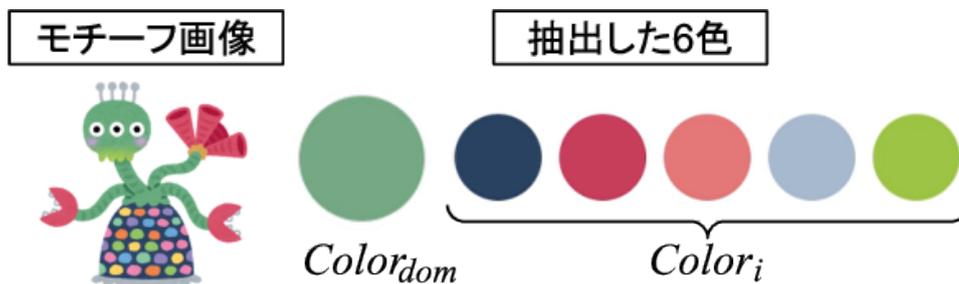


図 7.4: モチーフ画像から抽出した6色

領域 a, b, c, d にモチーフから抽出した色を適用するアルゴリズムについての説明する。まず、色の

¹<https://www.chifure.co.jp/beauty/lesson/makeup/eyeshadow.html>

²https://www.nomorerules.net/all_products/eye_shadow/

³<https://www.ai-teranagane.jp/logic/6659/>

⁴<https://mdpr.jp/beauty/detail/1877138>

扱いを簡単にするため、モチーフ画像をメディアンカット法を用いて6色に減色し、色を抽出する。メディアンカット法は原画像の画素のRGB値を、3次元の直方体の内にある座標として考え、その直方体を画素数が同じになるように中央値で分割していく手法である。減色によって取得した6色をRGB色空間からHSV色空間へと変換し、最も色面積が大きい色を支配色 $Color_{dom}$ とする。このとき、6色のうち、 $Color_{dom}$ 以外の5色を $Color_i (1 \leq i \leq 5)$ とする (図 7.4)。

図 7.3の領域 a は顔の上を覆う最も広い領域である。そのため、モチーフ画像に使われている最も色面積が大きい色である $Color_D$ を選択するのが最適であると考えた。このとき、領域 a の色を A とし、 $A = Color_{dom}$ とする。

図 7.3の領域 b は目の際に近い領域であり、領域 a よりも暗い色を使うのが一般的である。 $Color_i$ の明度を V_i とし、 A の明度を V_A としたとき、 $V_i < V_A$ となるすべての $Color_i$ を領域 b の色の候補とする。このとき選択された領域 b の色を B とする。

図 7.3の領域 c は領域 a, b を引き立たせるための色として反対色に近い色を選択する。反対色とは任意の色に対して、色相環上で正反対に位置する関係にある色のことを指す。 A, B の組み合わせに対する領域 c の色を C とし、候補を $Color_i$ から選択するとき、 A の色相を H_A とし、 B の色相を H_B とする。このとき、色相の範囲を $0 \leq h \leq 360$ とする。

$$H_{med} = \frac{(H_A + H_B)}{2} \quad (7.1)$$

として、 A, B の組み合わせに対する色相の中間点を求める。 $Color_i$ の色相を H_i とし、

$$H_{med} + 120 < H_i < H_{med} + 210 \quad (7.2)$$

の範囲に含まれる色相を持つ $Color_i$ を C の候補とする。 C として選択し得る候補が存在しない場合、領域 c は色を持たず 3 色のアイシャドウは生成されない。

図 7.3の領域 d は領域 a, b の上に重ねるハイライトとしての色を選択する。そのため、支配色よりも十分に明度が高い色を配置することにした。また、領域 c と領域 d が同じ色になることを避けるため、 C を選択しないこととする。領域 d の色を D とし、 $Color_{dom}$ と $Color_i$ の明度をそれぞれ V_{dom}, V_i としたとき、

$$Color_i \neq C \wedge V_{dom} + a < V_i \quad (7.3)$$

を満たす全ての $Color_i$ を D の候補として選択する。現時点の実装では $a = 0.1$ とした。また、 D として選択し得る候補が存在しない場合、領域 d は色を持たず 4 色のアイシャドウは生成されない。

7.2.2 アイライン



図 7.5: アイラインのタブの内容

アイラインは目の縁に沿って線を引くことで、目を大きく見せ、雰囲気を変化させる。特に目尻の描き方を変えることで、吊り目やタレ目に見せることが可能である。演じる役の雰囲気を演出するために、アイラインの目尻の描き方は複数の種類から選択できるべきだと考えた。また、色を変えることでアイシャドウと馴染ませたり、一方で、アイラインを強調することも可能である。

アイラインのタブの内容を図 7.5 に示す。タブ内の機能は大きく 2 つに分けられ、ユーザはアイラインの形状と色を選択することができる。タブの上部 (図 7.5-(a)) には、アイラインの形状の候補が 3 つ並ぶ。

タブの下部 (図 7.5-(b)) では、アイラインの色を選択できる。表示される色は、モチーフ画像の減色によって取得された 6 色に黒色を加えた 7 色とした。

目の縁は形状や大きさの個人差が最も大きい部位である。一方で、同じく目の縁に使用する、つけまつ毛のように、長過ぎる場合は、自身の目の大きさに合わせて切って使うことを想定されたメイク道具も存在する。現在の実装では、筆者の目の形状と大きさをモデルにアイラインを設計し、転写シールから切り抜く際にユーザ自身の目に合わせて調整してもらうことにした。後の実験において、転写シールの形状が被験者の目に合っていたかどうかを評価する。

7.2.3 アイペイント

アイペイントはモチーフ画像の一部をマスク画像として切り取り、マスク画像と模様画像をマスク形状に合わせて切り抜き合成することで生成する。モチーフ画像の一部をマスク画像として適用することで、モチーフの形状を的確に再現する。また、試作と同様にモチーフ画像の一部を模様画像と使用する

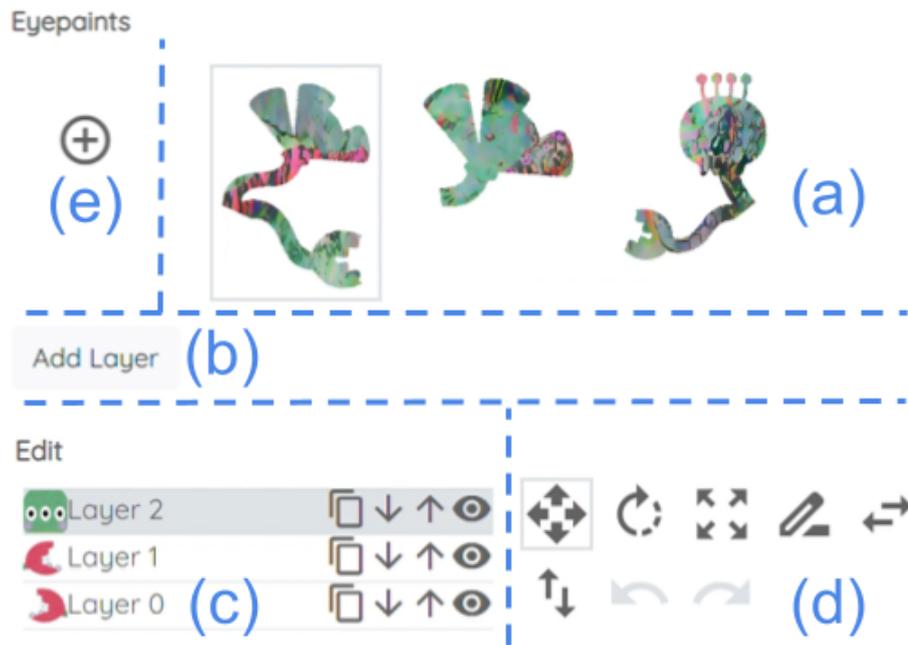


図 7.6: アイペイントのタブの内容

ことで、モチーフの色や柄を明確に表現する。さらに、別の手法により生成した模様画像を導入することで、より抽象的な表現も可能とした。

アイペイントのタブの内容を図 7.6 に示す。アイペイントの一覧 (図 7.6-(a)) から 1 つを選択し、“add layer”(図 7.6-(b)) ボタンを押すと、キャンバスにアイペイントが追加される。図 7.6-(c) には、追加されたアイペイントの一覧がレイヤー形式で表示されているレイヤーの右側にあるアイコンで、レイヤーの複製、下に移動、上に移動、可視・不可視の切り替えの操作を行える。図 7.6-(d) に並ぶアイコン群は、アイペイントをキャンバス上で編集するための機能である。左上から順に、移動、回転、拡大縮小、消しゴム、水平反転、垂直反転、編集を 1 つ戻る、編集を 1 つ進む機能である。

アイペイントのマスク画像生成には、ユーザが手動で生成する方法とシステムが自動的に生成する方法の 2 種類が存在する。ユーザが手動で生成する場合は、図 7.6-(e) を押すと専用のインターフェースが開き、編集後に図 7.6-(a) にアイペイントが追加される。自動作成によるマスク画像を使用したアイペイントは、モチーフ画像入力後、あらかじめ図 7.6-(a) に表示されている。また、どちらの生成手法においてもマスク画像に対する模様画像としてスタイル変換 [Gatys16] を用いた模様画像と合成を行う。手動で生成する場合は、さらに、ユーザがモチーフ画像から選択した領域を模様画像として合成し、合計で 2 つのアイペイントを作成する。



図 7.7: アイペイント作成インターフェース

アイペイントの手動生成

アイペイントを手動で作成するためのインターフェースを図 7.7に示す。図 7.7-(a) でアイペイントの形となるマスク画像の領域を選択し、図 7.7-(b) で模様画像となる領域を選択する。図 7.7-(c) には、マスク画像と模様画像として選択した領域を重ね合わせて合成した結果が表示される。出力されるアイペイントの大きさはマスク画像に合わせて、模様画像のサイズが調整される。

マスク画像の自動生成

通常のアイメイクでは目を大きく見せるために、アイシャドウやアイラインを目尻方向に延ばし、実際の目の大きさより長めに描く。アイペイントの形状を自動生成する際、モチーフ画像から細長く伸びた部分を抽出し、目尻方向に配置しやすい形状の作成を目指した。このため、画像領域から外に突出している部分をアイペイントの形状として用いることにした。

モチーフ画像を連結成分へ分割し、各成分について 3 個以下のマスク形状の候補を出力する。

はじめに、モチーフ画像が 512×512 ピクセルより大きい場合、画像を収まるサイズに縮小する。画像フォーマットが各チャンネル 1 バイトの RGBA であると仮定し、アルファチャンネルの値が 200 以上である点をモチーフ画像領域として二値化した。モチーフ画像領域から 8 連結成分を全て抽出し、連結成分の内部に含まれる穴もその成分に含まれるとした。ここで、8 連結成分とは、二値化画像中の 8 隣接で隣接している画素の集合のことである。本節では、各 8 連結成分の領域をモチーフ画像のパーツと呼

ぶ。抽出した全てのパーツの面積を計算し、面積が(全成分の合計面積) × 0.2 未満であるパーツを除外した。これは、小さいパーツはアイペイントの形状として適していないと考えたためである。

それぞれのパーツについて、画像からのボトルネック検出方法 [Wang11] を用いて、画像領域から外に飛び出している部分を検出した。この方法では、画像の境界上の、境界に沿った長さが大きく、ユークリッド距離が小さくなるような2点を選び、その2点をつなぐ線分で画像領域を切断する。成分の境界の長さを N とし、境界上を時計回りまたは反時計回りの任意の方向に一周するように点 p_1 から点 p_N を定義する。画像の境界上の2点 p_i と p_j について、次のスコアを計算した [Wang11]。

$$score(p_i, p_j) = \frac{D(p_i, p_j)}{\min\{L(p_i, p_j), L(p_j, p_i)\}} \quad (7.4)$$

ただし D はユークリッド距離、 L は境界に沿った距離である。 p_i と p_j をつなぐ線分が背景を横切った場合、そのような p_i と p_j は除外した。

式 (7.4) のスコアが小さくなる点のペア (i, j) を選び、それらの点をつなぐ線分で領域を切断し、面積の小さい方をマスク形状候補とする。ただし、単純にスコアが最小となるペアを複数個選ぶと、 i や j が僅かにしか異ならない点のペアが選ばれ、ほぼ同じ形状のマスク候補が複数個出力されてしまう。これを避けるため、次の操作を繰り返し、複数の点のペアを決定した。まず、式 (7.4) のスコアが最小となる2点 \hat{i}, \hat{j} を、

$$\hat{i}, \hat{j} = \operatorname{argmin}_{i, j} score(p_i, p_j) \quad (7.5)$$

として選んだ。 \hat{i} と \hat{j} をつなぐ線分で領域を切断し、面積の小さい方をマスク形状候補として出力した。

次に、 \hat{i}, \hat{j} の近傍の点

$$\begin{aligned} \hat{i} - aN \leq i \leq \hat{i} + aN \\ \hat{j} - aN \leq j \leq \hat{j} + aN \end{aligned} \quad (7.6)$$

となる点のペア (i, j) を除外した。また、現時点の実装では $a = 0.1$ としている。この操作を、出力されたマスク形状が上限 (3 個) に達するか、候補となる点のペアが無くなるまで繰り返した。

スタイル変換による模様づけ

モチーフ画像内に存在する柄や模様を、模様画像として使用する場合は、手動作成によって実現可能である。一方で、形状を保持しつつ画像内の柄や模様は使用せず、モチーフの雰囲気だけを表現したい場合がある。本実装ではモチーフ画像を抽象的に表現するために、スタイル変換を用いることにした。

スタイル変換とは、変換の対象となる元の画像 (コンテンツ画像) に、スタイルを与えるための参照画像 (スタイル画像) のテクスチャ、色調、筆触などの特徴を反映する技術である。今回の実装では、スタ

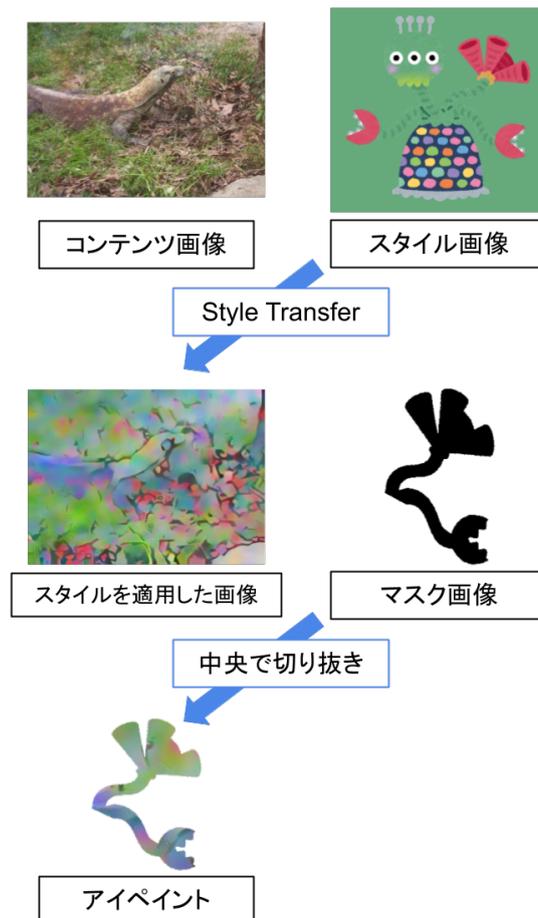


図 7.8: スタイル変換を利用したアイペイントへの模様付け

スタイル変換には Fast Style Transfer API⁵を利用した。本システムでスタイル変換を用いたときのアイペイントの作成方法を図 7.8に示す。スタイル変換を行うときには、スタイル画像とコンテンツ画像の 2 枚の画像が必要になる。

スタイル画像には、モチーフ画像を用いる。このとき、モチーフ画像には背景のない画像を用いているため、背景色が白色となり、スタイル変換の出力結果に白色が反映されてしまう。背景色の白色がアイペイントの模様として出力されるのを防ぐために、背景をモチーフ画像から抽出した支配色で塗りつぶした画像をスタイル画像とした。

多くのスタイル変換に関するモデルは自然画像をコンテンツ画像として学習している。そのため、コンテンツ画像には、ImageNet[Deng09]の画像を使用した。モチーフ画像を表す単語 1 語とコサイン類似度が最も高い単語を持つ ImageNet のラベルの画像群から、毎回ランダムに 1 枚画像を選出し、コンテンツ画像とすることでスタイル変換を行う。現在の実装では、モチーフ画像を表す単語 1 語は筆者が手動でラベル付けを行った。コサイン類似度の計算には、Wikipedia 2017, UMBC webbase コーパス、

⁵<https://deepai.org/machine-learning-model/fast-style-transfer>

statmt.org news データセットから学習した 100 万語の学習済み単語ベクトルを使用した [Mikolov18].

最後に、システムまたはユーザが作成したマスク画像の中心と、スタイル変換が適用された画像の中心を重ね合わせ、マスク画像の形に合わせて切り抜くことで、スタイル変換を施したアイペイントを生成した。

7.3 生成したデザイン 3 要素の印象評価

今回提案したモチーフの特徴を反映したアイメイクの要素である、アイシャドウ、アイライン、アイペイントの生成結果に対してアンケートによるデザイン評価実験を実施した。実験には、ゲンガー (ポケットモンスター)⁶、チョッパー (ONE PIECE)⁷、キュアバタフライ (ひろがるスカイ! プリキュア)⁸ の 3 種類のモチーフを使用した。使用モチーフは、アイシャドウの 4 色パレットが生成できるキャラクターに限定した。

回答者にはまず、モチーフの特徴を反映したアイメイクの例を提示してから、アンケートに回答してもらった。質問の内容は次の 3 項目である。

質問 1 図のアイシャドウの配色はまぶたの上のアイメイクとして良い配色だと思いますか? (36 枚の画像に対して)

質問 2 図に示した色のついた円形の個数は、左のモチーフを表現できていると思いますか? (15 枚の画像に対して)

質問 3 図のパーツは左のモチーフを抽象的に表現したアイペイントのパーツとして良いパーツだと思いますか? (36 枚の画像に対して)

それぞれの質問に対して、1 を「全くそう思わない」、5 を「とてもそう思う」として、5 段階のリッカートスケールで評価を行ってもらった。また、質問項目内の画像は、順序効果の影響を軽減するため、回答者ごとにランダムな順番で提示した。

質問 1 はアイシャドウの配色に関する質問であり、システムが提案したパレットの配色と、メディアカット法を用いて得た 6 色を重複なくランダムに配色したアイシャドウの画像を提示して評価してもらった。それぞれ、2 色、3 色、4 色のパレットを 3 種類のモチーフから 2 種類ずつ生成し、合計で 36 枚の画像を用意した。用意した画像の左側には、システムのキャンバス内に描画されるグラデーション表現のアイシャドウ画像を配置し、右側には、アイシャドウのタブに表示されるパレット画像を配置した。

⁶<https://zukan.pokemon.co.jp/zukan-api/up/images/\index/72c82f8be362d1b53ae308d706728411.png>

⁷https://toy.bandai.co.jp/assets/tamagotchi/\images/chopper/img_chara01.png

⁸https://www.toei-anim.co.jp/tv/precure/assets/img/\character/p_03_style1-2_sp.png

質問2はアイラインの色の選択肢の個数に関する質問であり、メディアンカット法を用いて3種類のモチーフから2色、4色、6色、8色、10色を抽出し、合計で15枚の画像を用意した。用意した画像の左側にはモチーフの画像を配置し、右側にはアイラインのタブに表示される色の選択肢と同様に、抽出した色を円形で表現した画像を配置した。

質問3はアイペイントの形状と模様に関する質問である。アイペイントの設問に回答してもらう前に、アイペイントについて説明を記載し使用例を提示した。形状について、システムが提案した3種類のマスク画像、および、5.3.2節で提案したアルゴリズムのうち i, j をランダムに決定した3種類のマスク画像を、3種類のモチーフから合計18種類生成した。模様については、スタイル変換を用いた模様画像と、各ピクセルに対してモチーフから抽出した6色をランダムに選択して着色した模様画像の2種類を生成した。形状と模様をかけ合わせると、合計で36枚の画像を用意した。用意した画像の左側にはモチーフの画像を配置し、右側には形状と模様を合成したアイペイントの図を配置した。今回の実験では、スタイル変換で使用するコンテンツ画像はモチーフごとに差が出ないように、ImageNetから選択した1枚の画像をすべてのアイペイントに適用し、モデルはarbitrary-image-stylization-v1-256⁹を使用した。

7.3.1 結果

アイシャドウ、アイラインの化粧品を使用した経験がある35名に、アンケートに回答してもらった。回答者は全員10代から30代の女性であり、アンケートはweb上で実施した。

アイシャドウ

表 7.1: アイシャドウ配色評価結果

	平均値	標準誤差	中央値
iMake	3.77	0.10	3.78
ランダム配色	3.13	0.10	2.94

質問1の評価結果を表7.1と図7.9に示す。表7.1に示す数値は、ユーザごとにシステムとランダムの配色に対する評価の平均値を算出し、ユーザごとの平均値に対して、平均値、標準誤差、中央値を求めた結果である。平均値、中央値ともに、システムによる配色が高い数値を示した。

加えて、t検定における片側検定を行い、有意差があるかを検証した。その結果、p値は $1.81e-11$ であり、システムの配色とランダムの配色の評価に有意に差があることが認められた。

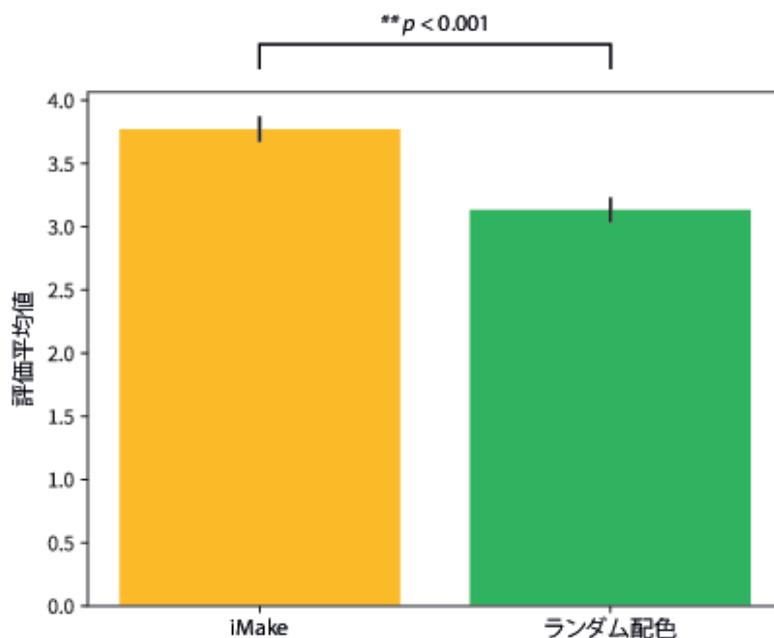


図 7.9: 配色に対する回答者ごとの評価平均値

表 7.2: アイラインの色数評価結果 (数値)

	平均値	標準誤差	中央値
2 色	3.15	0.13	3.33
4 色	3.58	0.14	3.67
6 色	3.60	0.12	3.67
8 色	3.54	0.16	3.67
10 色	3.47	0.21	3.67

アイライン

質問 2 の評価結果を表 7.2 と図 7.10 に示す。表 7.2 に示す数値は、ユーザごとに 2 色、4 色、6 色、8 色、10 色に対する評価の平均値を算出し、ユーザごとの平均値に対して、平均値、標準誤差、中央値を求めた結果である。平均値は 6 色が最も高く、中央値は 2 色以外は変わらない結果となった。

また、本システムで用いた色数である 6 色と、その他の色数に対して、t 検定における片側検定を行い、有意差があるかを検証した。その結果、2 色に対しては Bonferroni 補正による多重検定の有意水準の補正後の P 値が $P = 0.03$ となり有意差が認められたが、その他の色数では認められなかった。

アイペイント

質問 3 の評価結果を表 7.3 と図 7.11 に示す。表 7.3 に示す数値は、システムが生成した形状と i, j をランダムに決定した形状、および、スタイル変換を用いた模様と、各ピクセルごとにランダムに色を決定

⁹<https://tfhub.dev/google/magenta/arbitrary-image-stylization-v1-256/2>

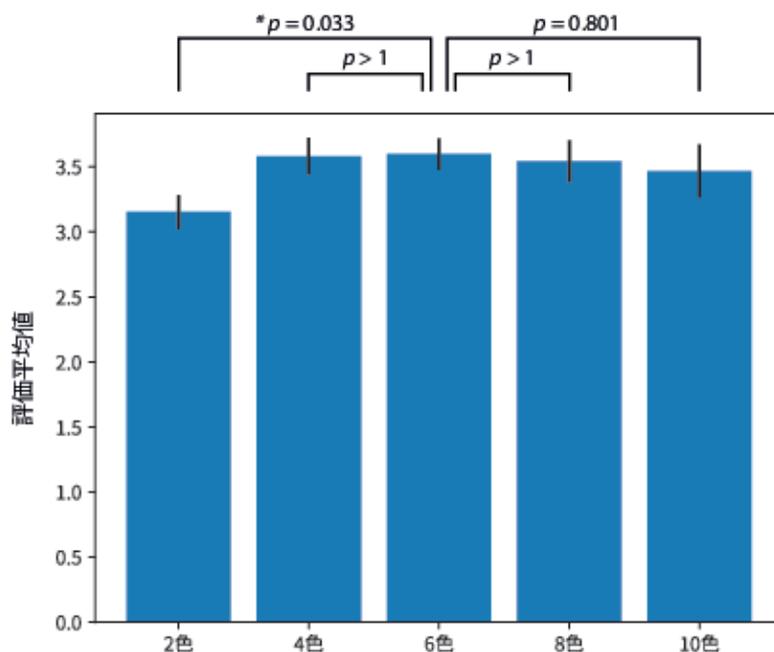


図 7.10: アイラインの色数評価結果 (グラフ)

表 7.3: アイペイント評価結果

	平均値	標準誤差	中央値
システム形状	2.63	0.10	2.56
ランダム形状	2.34	0.11	2.39
スタイル変換	2.43	0.11	2.33
ランダム模様	2.53	0.11	2.44

した模様に対する評価の平均値をユーザごとに算出し、ユーザごとの平均値に対して、平均値、標準誤差、中央値を求めた結果である。平均値ではシステムが生成した形状が、 i, j をランダムに決定した形状よりも高い数値を示した。一方で、スタイル変換を用いた模様の平均値は、各ピクセルごとにランダムに色を決定した模様よりも低い結果となった。

加えて、形状と模様について、二要因分散分析による検定を行い、有意差があるかを検証した。その結果、形状の主効果においてはP値は $2.14e-07$ であり、システムが生成した形状とランダムに生成された形状の評価に有意に差があることが認められた。一方で、模様的主効果においてはP値は $1.82e-01$ であり、スタイル変換を用いた模様とランダムに生成した模様の評価に有意差は認められなかった。交互作用についてもP値は $3.48e-01$ であり、有意差は認められなかったため、形状の評価値の差は模様依存しないといえる。

7.3.2 考察と議論

実験の結果より、アイシャドウ、アイライン、アイペイントのデザインに対する考察を行う。

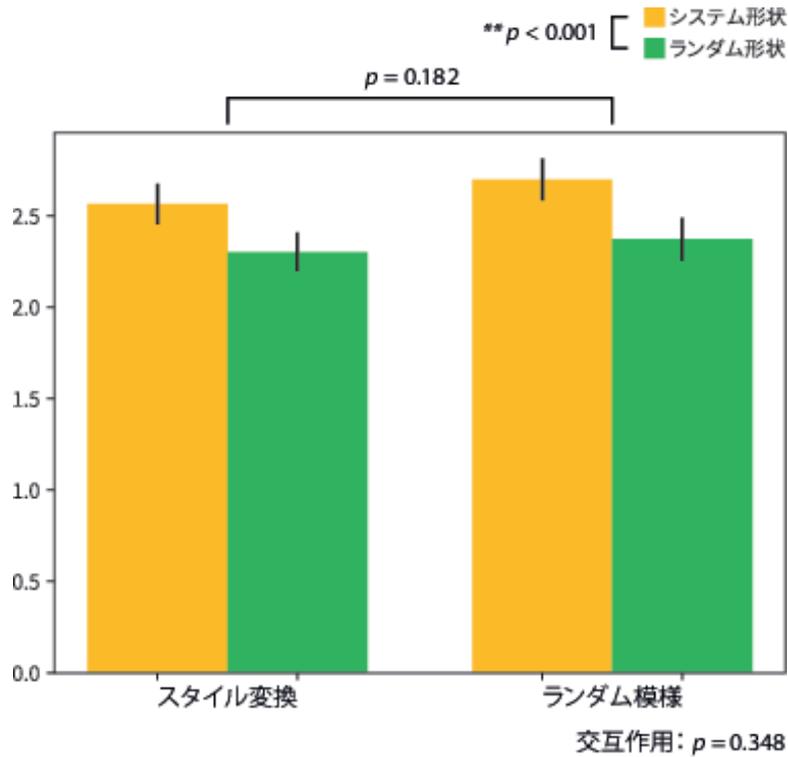


図 7.11: アイペイントの形状と模様に対する回答者ごとの評価平均値

アイシャドウに関して、本論文で提案したアルゴリズムは、ランダムな配色と比較して有意に良い評価を得られた。アイシャドウの配色には一定の規則性が存在し、iMake でユーザのデザイン作成補助をする項目として妥当であったと言える。現在のアルゴリズムでは彩度を考慮していないため、今後は彩度を組み込むことで、より良いアルゴリズムの開発を行っていく。

アイラインの色数に関して、平均値は現在採用している 6 色が最も良い評価を得たが、4 色、8 色、10 色との有意差は認められなかった。今後はモチーフから抽出する色の個数をユーザが選択できるような機能を組み込むと、デザインのバリエーションがより豊かになると考えられる。

アイペイントに関して、形状については、本論文で提案したアルゴリズムが有意に良い評価を得られた。一方で、すべての項目において中央値は 2 を示しており、アイシャドウやアイラインと比較すると全体的に低い評価となった。モチーフから形状を切り出す際、特に人間のモチーフに今回のアルゴリズムを適用すると、髪の毛の端や手などがマスク画像として生成される点が問題点として挙げられる。これは、モチーフ画像から外に突出している部分を形状として用いるためである。今後、改良していく上で、モチーフの輪郭を見るだけでなく、境界検出などによりモチーフ画像をより意味のある領域に分割すると、直感と一致した形状を切り出すことができる可能性がある。また、模様についても、スタイル変換とランダム生成の間に有意差は認められず、どのような模様がアイペイントとして適しているのか、さらに模索する必要がある。

7.4 従来の化粧方法とシステムを用いた化粧方法による化粧過程の比較実験



図 7.12: 実験に用いた 2 種類のモチーフ画像 (出典: いらすとや)

表 7.4: システム評価実験の作業後アンケート内容

2 択設問
[質問 1] デザインを考えやすかったのはどちらですか？ (iMake・従来のメイク)
[質問 2] メイク (肌に対して施す過程) が簡単だったのはどちらですか？ (iMake・従来のメイク)
[質問 3] iMake に表示されていたアイシャドウの色は適切でしたか？ (はい・いいえ)
[質問 4] iMake に表示されていたアイラインの種類と色は適切でしたか？ (はい・いいえ)
[質問 5] iMake に表示されていたアイペイントは適切でしたか？ (はい・いいえ)
[質問 6] 転写シールは目の形に合っていましたか？ (はい・いいえ)
[質問 7] 転写シールの貼り付けは簡単でしたか？ (はい・いいえ)
[質問 8] モチーフの特徴を反映したアイメイクとしての出来が良かったのはどちらですか？ (iMake・従来のメイク)
[質問 9] メイクが完了した時の満足度が高かったのはどちらですか？ (iMake・従来のメイク)
自由記述
従来のメイクの良かった点を挙げてください
従来のメイクの改善点を挙げてください
iMake の良かった点を挙げてください
iMake の改善点を挙げてください
実験の内容について自由に感想を述べてください

本システムの目的は従来の化粧方法でモチーフの特徴を反映したアイメイクをする際に課題となる 3 つの項目を解決することである。

課題 1 メイクの技量によって仕上がりに個人差がある

課題 2 所持している化粧品によって使用できる色が制限される

課題 3 モチーフの色彩や形状の表現を反映した独創性の高いデザインを考案する技量に個人差がある

それぞれの課題に対して、課題1を解決するために転写シールを使用することでメイクの技量による仕上がりの差をなくす。課題2を解決するために、インクジェットプリンタの使用によって、CMYKで出力可能な色をすべて使えるようにする。課題3を解決するために、モチーフの特徴を反映したアイメイクをアイシャドウ、アイライン、アイペイントの3つの要素に分解し、システムがそれぞれのデザイン候補を提案することでユーザのデザイン作成を補助する。

提案したシステムで、これらの課題が解決できているかを調査するため、被験者に従来のメイク方法とiMakeを使用したアイメイクの両方を実施してもらい、その結果を比較する実験を行った。モチーフは従来のメイク方法とiMakeでそれぞれ使用する2種類の画像をmonster, kirinとして用意し(図7.12), 同一の被験者に対して従来のメイク方法とiMakeで同じモチーフは使用しなかった。順序効果の影響を軽減するため、従来のメイク方法とiMakeを実施する順序、および、それぞれのアイメイクで使用する2種類のモチーフを入れ替えた4種類の組み合わせに関して、被験者数を均等にした。実験は筆者の自宅で行い、化粧をする際は洗面所にある三面鏡を使用した。

最初に、被験者に対して、モチーフに基づいたデザインを考えてアイメイクを施すアイメイクについて説明し、一例を提示した。

従来のメイク方法によるアイメイクを行う場合、被験者には紙に印刷したモチーフ画像を提示し、モチーフをもとにデザインを考え、アイシャドウとペンシルタイプのアイライナーを用いてメイクを実施するよう指示した。実験においては、被験者が所持している化粧品によってアイメイクが制限されることを防ぐため、アイシャドウ50色とアイライナー12色を用意した。

iMakeによるアイメイクを行う場合、まず被験者にシステムの使用方法を説明した。その後、実験用に用意した2種類のモチーフとは別のモチーフ画像を使用して、被験者にアプリケーションを実際に触ってもらいながら、分からない部分について質問を受け付けた。被験者からの質問が終了した後、紙に印刷したモチーフ画像を提示し、アプリケーション上で同様の画像を読み込み、アイメイクのデザインを開始してもらった。完成した後、転写シールに出来上がったアイメイクのデザインを印刷し、被験者に使用方法を説明した。説明中に、使用方法を確認してもらうため、一度だけ腕の内側に転写シールを貼り付けてもらった。その後、被験者に転写シールからデザインしたアイメイクを切り抜いてもらい、練習等はせずに顔に貼り付けてもらった。

作業後のアンケートの内容を表7.4に示す。2択質問のうち[質問3]-[質問6]については、いいえと回答した場合、その理由を自由記述で述べてもらった。

本実験において、iMakeでは、使用するモチーフが決まっているため、システムが行うアイペイントのマスク画像の切り出しは予め行い、アプリケーションに組み込んだ。また、アイペイントをスタイル変換

表 7.5: 作業時間の結果

	平均値	標準誤差	中央値
(a) iMake	823.00 秒	121.70 秒	780.00 秒
(b) 従来のメイク	683.88 秒	122.13 秒	614.00 秒
(c) デザイン	420.63 秒	111.20 秒	359.50 秒
(d) 貼り付け	402.38 秒	29.17 秒	420.50 秒

を用いて作成するためのコンテンツ画像を ImageNet から選出するため、用意した 2 種類のモチーフ画像をそれぞれ monster, kirin としてコサイン類似度を算出した。その結果、最も類似度の近い ImageNet のラベルは、monster が“Gila monster”で 1.0, kirin が“Komodo dragon”で 0.5 となった。両ラベルにはそれぞれ 1300 枚の画像が紐付けられており、アイペイントを作成する際には、これらの画像から毎回ランダムに 1 枚選出しコンテンツ画像として使用した。

実験においては時間制限を設けず、被験者に自由にデザインと化粧を行ってもらった。化粧中の作業時間を計測し、従来のメイクと iMake での化粧が終わったあと、アンケートに回答してもらった。このとき、iMake を用いた作業時間は、アプリケーションを用いてデザインをした時間と転写シールからアイメイクを切り抜いて瞼に貼り付けるまでの時間を合算したものとする。

7.4.1 結果



図 7.13: (左) 従来のメイク方法による化粧, (右) iMake による化粧をした結果

アイシャドウ, アイラインの化粧品を使用した経験がある被験者 8 名に実験に参加してもらった。被験者は 8 名全員が 20 代から 30 代の女性であった。図 7.13 に従来のメイク方法と, iMake を用いて化粧をした結果を示す。同一の被験者が左右で対応するように図は配置している。

従来のメイク方法の作業時間と iMake を用いた作業時間を比較した結果を表 7.5 と図 7.14 に示す。図 7.14 の P 値は iMake の作業時間 (表 7.5-(a)) と従来のメイク方法の作業時間 (表 7.5-(b)) に対して t 検定

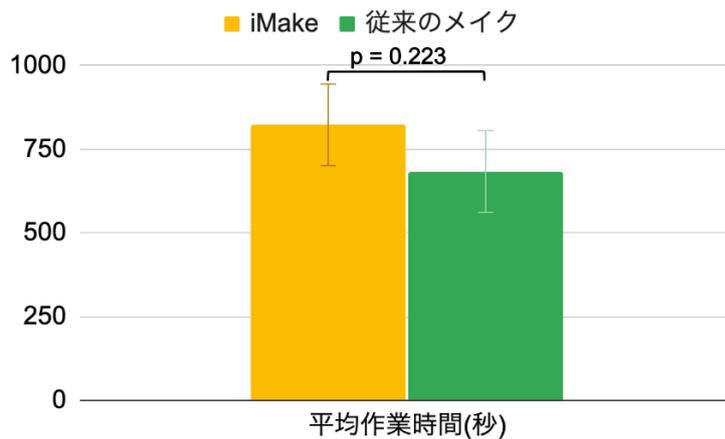


図 7.14: iMake および従来のメイクの平均作業時間 (秒)

を用いた結果を示している。iMake の作業時間の内訳として、アプリケーションを用いてデザインした作業時間を表 7.5-(c) に、転写シールを顔に貼り付けるための作業時間を表 7.5-(d) に示す。

次に各課題について、設問および回答結果に関する報告を行い、また自由記述による回答についても、著者が各課題ごとに分類した結果を記述していく。

課題 1

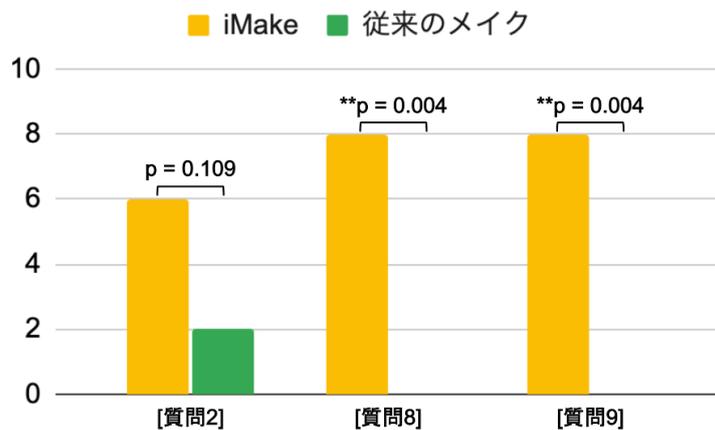


図 7.15: 課題 1 に関する設問の結果

課題 1 と関連する設問は以下の 3 つである。

[質問 2] メイク (肌に対して施す過程) が簡単だったのはどちらですか？ (iMake・従来のメイク)

[質問 8] モチーフの特徴を反映したアイメイクとしての出来が良かったのはどちらですか？ (iMake・従来のメイク)

表 7.6: 課題 1 に関する自由記述の回答

従来のメイクの良かった点
<ul style="list-style-type: none"> ・慣れているためメイクする作業そのものはやりやすかった ・一般的なメイク道具を使用する点
従来のメイクの改善点
<ul style="list-style-type: none"> ・複雑なメイクは当人の技術や器用さに依存するため、メイクに慣れていない人には難しいと感じた ・想定していた顔の部分にうまくアイシャドウやアイラインをのせられなかったり、想定していたデザインにたどり着けないことがあると感じた ・鏡越しデザインに慣れていないと難しい
iMake の良かった点
<ul style="list-style-type: none"> ・高いデザイン性が再現されやすい ・シールゆえ崩れにくいので、細かい複雑なデザインでも長時間キープしやすい ・自分の手書きだと書ききれない細かい表現ができるのがよかった ・想像以上に手軽にシールを貼ることができて驚いた ・メイク自体がとても楽で、特にアイシャドウがきれいに塗れたのがよかった
iMake の改善点
<ul style="list-style-type: none"> ・貼り付け時に台紙を裏返すため、デザインの位置が確認しづらい (2 人) ・数回の練習が必要、慣れるともっと素早くできそうだと感じた

[質問 9] メイクが完了した時の満足度が高かったのはどちらですか？ (iMake・従来のメイク)

これらの設問を被験者に評価してもらうことで、提案システムが従来のメイク方法と比較して、簡単にメイクができるかどうか、仕上がりに個人差があるモチーフの特徴を反映したアイメイクにおいて一定の水準を担保できるかどうかを検証する。

課題 1 の設問に対する結果を図 7.15 に示す。図 7.15 の P 値は、各設問に対して発生率 0.5 を帰無仮説として二項検定における片側検定を行った結果を示している。[質問 2] において、6 人が「iMake」を選択し、2 人が「従来のメイク」を選択した。[質問 8] において、8 人全員が「iMake」を選択した。[質問 9] において、8 人全員が「iMake」を選択した。

自由記述による回答をまとめた結果を表 7.6 に示す。

課題 2

本実験においては、被験者が所持している化粧品によってアイメイクアートが制限されることを防ぐため、あらかじめ、アイシャドウ 50 色とアイライナー 12 色を用意した。また、個人が所有する化粧品の色の種類と比較して、プリンタが出力できる色の種類が多いのは明らかである。そのため、設問として 2 択で評価することはせず、自由記述の設問のみとした。自由記述による回答をまとめた結果を表 7.7 に示す。

表 7.7: 課題 2 に関する自由記述の回答

従来のメイクの良かった点
<ul style="list-style-type: none"> ・ ラメなど立体感のあるアイシャドウがあった (2 人) ・ 普段使わない色だったのでわくわくがあった
従来のメイクの改善点
<ul style="list-style-type: none"> ・ 発色が悪い (2 人) ・ アイシャドウやアイラインを重ねると滲んだり色が混ざったりしてしまう (2 人) ・ 色を重ね塗できないので、思った色味ならなかったりする ・ 見た目の色と実際に肌に乗せたときの色が違う
iMake の良かった点
<ul style="list-style-type: none"> ・ 発色が良い ・ 色の感じや発色が思い通りの仕上がりになる (2 人)
iMake の改善点
<ul style="list-style-type: none"> ・ ラメなどの立体感を出すのは難しいと感じた ・ プリンタによって色の出方が変わるところ

課題 3

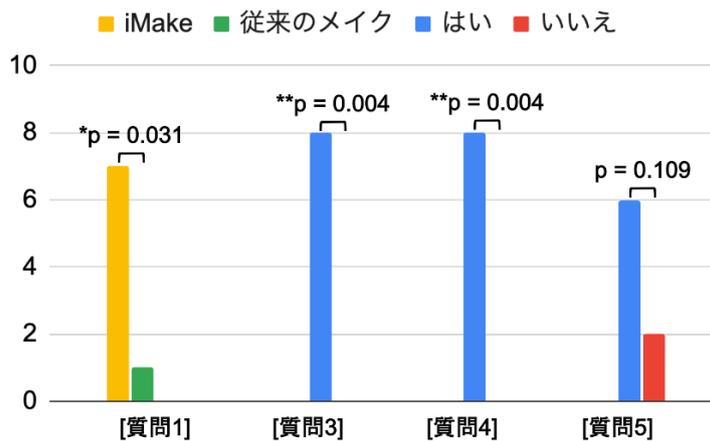


図 7.16: 課題 3 に関する設問の結果

課題 3 と関連する設問は表 7.4 のうち、[質問 1]、[質問 3]、[質問 4]、[質問 5] である。課題 3 と関連する設問は以下の 4 つである。

[質問 1] デザインを考えやすかったのはどちらですか？ (iMake・従来のメイク)

[質問 3] iMake に表示されていたアイシャドウの色は適切でしたか？ (はい・いいえ)

[質問 4] iMake に表示されていたアイラインの種類と色は適切でしたか？ (はい・いいえ)

[質問 5] iMake に表示されていたアイペイントは適切でしたか？ (はい・いいえ)

表 7.8: 課題 3 に関する自由記述の回答

従来のメイクの良かった点
<ul style="list-style-type: none"> ・目の形に応じてグラデーションやラインをデザインできる(2人) ・自分の描きたい場所を確認しやすい ・モチーフからデザインへの落としこみが最初の時点ではふわっとしていても、メイクしながら考えられる(2人)
従来のメイクの改善点
<ul style="list-style-type: none"> ・デザインのアイデアが思い浮かばなかった ・デモがないため、一発勝負で失敗できない ・デザインの自由度が高いゆえにどうすればいいか戸惑った ・デザインセンスや器用さがないと描けない(2人) ・デザインを考えながらメイクを同時に行うことは難しいと感じた
iMake の良かった点
<ul style="list-style-type: none"> ・いろんなデザインを事前にシミュレーションできる(2人) ・パソコン上でデザインするので皮膚への負担がないまま、デザインのやり直しや微調整が何度でも自由にできる(3人) ・デザインが不慣れな人でも、モチーフを彷彿とさせるアイメイクが簡単にできる ・デザインをリコメンドしてもらえるのがよかった。リコメンドを見てさらなるアイデアを思いつくなど、アイメイクの創作の幅が広がりそうだと思った(2人) ・リコメンドに加えて、アイメイクのデザインに慣れている人が自由にカスタマイズできるようにもなっているので、デザイン初心者もプロも使える良いシステムだと思った ・自分がデザインしているという主体感を残しつつ、システムの提案を取り入れて試行錯誤できる点
iMake の改善点
<ul style="list-style-type: none"> ・消しゴムで消すだけだと難しいので、細かい部分を付け足して描いたりできる(3人) ・画像切り抜きをカーブなどに対応してもらえるとデザインしやすい ・マスク機能が使いにくい ・消しゴムの太さを変えたい ・アイペイントを作るのがもう少し簡単だ(2人)

これらの設問を被験者に評価してもらうことで、提案システムがアイメイクのデザイン作成を補助できているかどうかを検証する。

課題 3 の設問に対する結果を図 7.16 に示す。図 7.16 の P 値は、各設問に対して発生率 0.5 を帰無仮説として二項検定における片側検定を行った結果を示している。[質問 1] において、7 人が「iMake」を選択し、1 人が「従来のメイク」を選択した。[質問 3] において、8 人が「はい」を選択し、0 人が「いいえ」を選択した。[質問 4] において、8 人が「はい」を選択し、0 人が「いいえ」を選択した。[質問 5] において、6 人が「はい」を選択し、2 人が「いいえ」を選択した。[質問 5] について、いいえと回答した理由について「細かい切り抜きが困難だった」「アーティスティックなデザインの使いどころが難しかった。自分でデザインするときのマスク処理が使いにくかった」という理由が挙がった。

自由記述による回答をまとめた結果を表 7.8 に示す。

その他

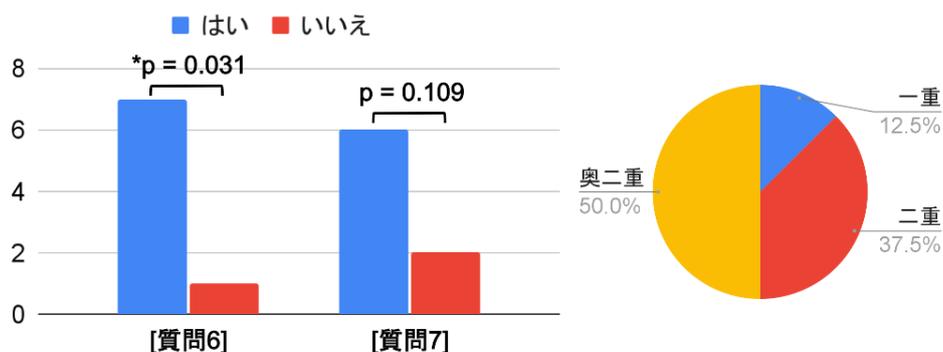


図 7.17: (左) 転写シールに関する設問の結果, (右) 目の形に関する設問の結果

表 7.9: 実験の感想

- ・転写シールが想像以上に楽かつ綺麗でびっくりした
- ・メイクオフがしやすいという点も実用性があると感じた
- ・従来のメイクで一からデザインする工程と、実際にメイク道具で着色する工程は両方とも時間がかかり、難しかった。デザインとメイクの両方を支援する提案システムは有用であると感じた
- ・実際に転写シールを貼ったときは、とってもわくわくしてたくさん写真を撮ってしまった。デジタル上で仕上がりを見たときよりも、よりテンションが上がった
- ・メイクの習熟度、例えば、普段どれくらいメイクをするか、普段のメイクがどれくらい複雑か、派手なデザインメイクをしたことがあるか、などにもよって、iMake への印象が異なりそうだなと感じた

先に挙げた 3 つの課題とは別に、転写シールに関する設問に 2 択で回答してもらった。関連する設問は以下の 2 つである。

[質問 6] 転写シールは目の形に合っていましたか? (はい・いいえ)

[質問 7] 転写シールの貼り付けは簡単でしたか? (はい・いいえ)

これらの設問に対する結果を図 7.17-(左) に示す。図 7.17 の P 値は、各設問に対して発生率 0.5 を帰無仮説として二項検定における片側検定を行った結果を示している。[質問 6] において、7 人が「はい」を選択し、1 人が「いいえ」を選択した。いいえ、と回答した理由については「奥二重のため、アイラインが二重幅に持っていかれた」という理由が挙げられた。[質問 7] において、6 人が「はい」を選択し、2 人が「いいえ」を選択した。

表 7.4に加えて、後日、被験者の目の形について一重、二重、奥二重から回答をしてもらった。その結果を図 7.17-(右) に示す。被験者のうち 1 人が「一重」と回答し、3 人が「二重」と回答し、4 人が「奥二重」と回答した。

実験の感想をまとめた結果を表 7.9に示す。

7.4.2 考察と議論

実験の結果より、iMake と従来のメイク方法の作業時間には明確な差が無いことが分かった。一方で、課題 1 の自由記述のコメントにあるように、被験者は従来のメイク方法に慣れていると予想できる。転写シールを瞼に貼り付けるという作業に対して、練習して慣れれば iMake での作業時間が減る可能性が十分に有り得る。

次に、課題ごとに結果について考察を行う。課題 1 において、iMake の方がモチーフの特徴を反映したアイメイクとしての出来が良く、被験者の満足度も高いことが示された。この結果から、転写シールを用いることで、ユーザのメイク技量による仕上がりの差を無くすことが可能であることが示唆された。そのため、提案システムはモチーフの特徴を反映したアイメイクとして一定の水準を保つことができると言える。このように、課題 1 の解決手段として転写シールを用いることは有用であると考えられる。一方で、転写シールを瞼に貼り付ける過程については、慣れが必要であることが分かった。先述のように、メイク道具に関する慣れは作業時間に影響を与える。また、貼り付け時に裏返す必要があるため、デザインを確認するために何度か表に返して見る必要があり、作業時間が増加してしまう場面も見受けられた。このような課題を解決するためには、瞼に貼り付ける作業を補助するような工夫が必要である。例えば、転写シールの裏側に貼り付け時のガイドを印刷するなどが考えられる。

課題 2 において、インクジェットプリンタは個人が所有する化粧品よりも多くの色を出力できる。所有する化粧品によってアイメイクが制限されてしまう課題に対して、インクジェットプリンタの使用は有用である。一方で、ラメなどのキラキラした質感を再現できない点や、プリンタによって色の出力が変わってしまう問題があることが分かった。ラメの質感については、今後、システム上でアイシャドウの部分にラメのテクスチャを取り入れることで質感を再現できるかどうかを検討していきたい。また、プリンタによって色の出力が変わる点については、複数のプリンタを用意し、実際に出力した色がどのように変化するかを実験する必要がある。

課題 3 において、提案システムが適切にアイシャドウ、アイラインの候補を提示できていることが分かった。また、iMake の方がデザインを作成しやすいことが分かった。提案システムはユーザのデザイン作成を補助し、課題 3 の解決手段となり得ることが示された。しかしながら、アイペイントの候補の

提示に関しては改善の余地がある。矩形での切り抜きや不要な部分の削除をするツールだけではアイペイントの形状を作成するのに不十分である。カーブによる切り抜きや細かい部分の付け足しなどの機能拡張が必要であるとともに、現状の機能の使いにくさやアイペイント作成の手順の煩雑さを解消するようなインタフェースの改善が必要である。現在の実装では、アイペイントについてはユーザが編集可能な機能が多かった。そのため、今後はアイシャドウやアイラインのように、おおまかなデザインを選択するだけで、おおよそのアイペイントの位置や向きの候補を提示してくれるようなシステムへ改良することを検討する。

最後に、転写シールが目の形に合っているかどうかについて、合っていると回答した被験者が多かった。この結果から、個人の目の大きさや幅の違いは転写シールを使用したアイメイクにおいて大きな問題にはならないと考えられる。しかしながら、課題3の自由記述のコメントに「目の形に応じてグラデーションやラインをデザインできるところ(2人)」を従来のメイクの良い点として挙げている被験者がいる。さらに、転写シールが合わなかったと回答した被験者は奥二重だったが、全被験者(8人)のうち4人が奥二重であると回答しており、他の3人は合っていたと回答した。転写シールそのもののサイズについては、個人差が大きく出ることはないが、目の形や特徴を考慮した方がより洗練されたデザインになることが示唆された。特にアイラインはユーザの目の形状を解析し、それぞれの目の形に合うアイラインのスタイルを推薦するシステムも存在する[Wang22]。このように、iMake上でもユーザの目の形に合ったスタイルをデザインに反映することで、より個人に合ったアイメイクを提供することが期待できる。

7.5 従来の化粧とシステムを用いた化粧の印象評価

7.4節の実験で得た、従来のメイク方法と転写シールを用いた化粧の結果画像に対して、

- モチーフがアイメイクデザインに反映できているか
- 通常のメイクと舞台メイクのどちらに近い化粧デザインか

を28名に評価してもらった。

設問は以下の2点である。

[質問1] 左図のモチーフの色や形状の特徴を右図のアイメイクデザインが表現できているかどうかを5段階で評価してください(1: 全く表現できていない, 5: とても表現できている)

[質問2] 左図は通常のメイク写真, 右図は舞台メイクの写真です。中央の図がどちらに近いかを5段階で評価してください(1: 通常のメイク, 5: 舞台メイク)

7.4節の実験で得た、従来メイク方法による化粧をしたメイクアップ画像8枚、iMakeを用いてメイクアップした画像8枚、合計16枚の画像をそれぞれの質問で評価してもらった。

質問1では、モチーフを左に配置し、7.4節の実験で得た被験者のメイクアップ画像を右に配置した図を提示し、評価をしてもらった。質問2は、舞台メイクは通常の化粧と比較して、アイシャドウが濃く、アイラインを目の形を大幅に越えて引くのが特徴であることを伝えた上で回答してもらった。左に通常のメイク写真の例を4種類配置、中央に被験者のメイクアップ画像を配置、右に舞台メイクの写真の例を4種類配置した画像を提示した。

また、順序効果の影響を軽減するため、質問は回答者ごとにランダムに提示した。

7.5.1 結果

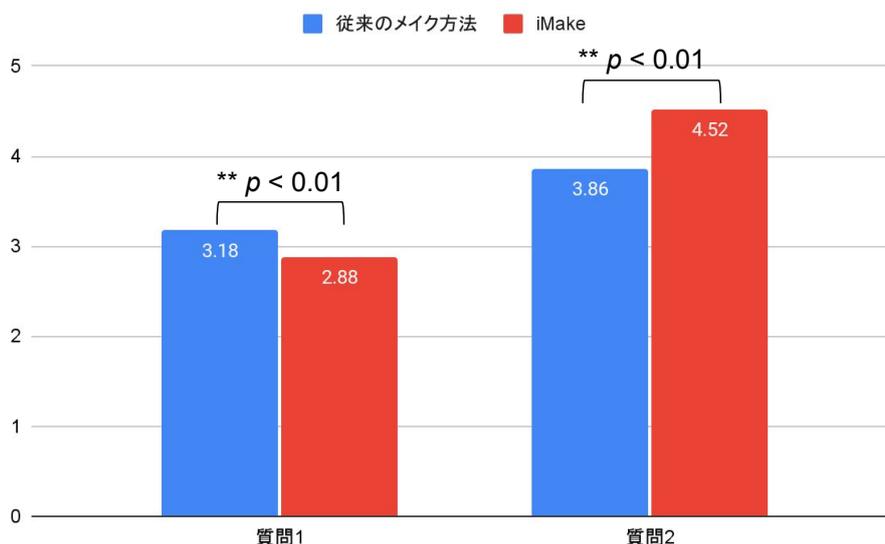


図 7.18: 質問1と質問2の平均値

図 7.18にそれぞれの設問における平均値の結果を示す。質問1において、従来メイク方法の平均値は3.18、中央値は3.00、標準偏差は1.18だった。また、iMakeを用いたメイク方法の平均値は2.88、中央値は3.00、標準偏差は1.00だった。t検定による片側検定の結果、p値は2.01e-03となり、従来メイク方法を用いた化粧の方がモチーフを表現できるという点において有意差が認められた。

質問2において、従来メイク方法の平均値は3.86、中央値は4.00、標準偏差は1.35だった。また、iMakeを用いたメイク方法の平均値は4.52、中央値は5.00、標準偏差は0.64だった。t検定による片側検定の結果、p値は6.78e-11となり、iMakeを用いた化粧の方が舞台メイクに近いという点において有意差が認められた。

7.5.2 考察と議論

質問1において、従来のメイク方法を用いた化粧の方がモチーフを表現できているという結果になった。一方で、7.4節で行った実験においても、「モチーフの特徴を反映したアイメイクとしての出来が良かったのはどちらですか?」という設問を設けており、iMakeの方が良かったという結果を得ている。これは、自身でデザインしたアイメイクデザインと、他者から見たアイメイクデザインでは評価に差が出ることを示唆している。舞台において、役者は観客に見られるものなので、デザインに対する評価の差を縮めていく必要がある。

一方で、質問2においては、iMakeを用いた化粧の方が舞台メイクに近いという結果になった。また、標準偏差についても、iMakeを用いた化粧の方がばらつきが小さい結果であった。これは、iMakeを用いた化粧は安定して舞台メイクに近いメイクデザインの作成が可能であることを示唆している。

今後は実際の舞台上でiMakeを用いた化粧をした結果、観客からどのように見えるかなどの評価を行いたい。

7.6 まとめ

本章では、プロトタイプ実装を改良し、3つの実験を行い、その評価と結果を示した。7.2節で、改良したシステムの内容を示した。7.3節では、提案システムが生成したアイシャドウ、アイライン、アイペイントのデザイン評価を行った内容と、その結果、考察を示した。7.4節では、提案システムと従来の化粧方法によるアイメイクの比較実験を行った内容と、その結果、考察を示した。6.2節で挙げた3つの課題において、課題1では、iMakeはユーザのメイク技量による仕上がりの差を無くしモチーフの特徴を反映したアイメイクとして一定の水準を保つことができた。課題2では、インクジェットプリンタを使用したアイメイクに対する可能性が示唆されたが、ラメに関する課題が残された。最後に、課題3では、iMakeがモチーフの特徴を反映したアイメイクのデザイン作成において有用であることが示されたが、アイペイントの作成には改善の余地があることが明らかになった。

7.4節の実験で得た化粧画像を用いた第三者による評価実験の内容と結果を7.5節で示した。従来のメイク方法の方がモチーフの特徴が表現できているという結果になったものの、7.4節ではiMakeの方がモチーフの特徴を表現できているという結果だったため、化粧のデザインを行った自身と他者から見たデザインの評価差を埋めていく必要がある。また、iMakeの方が舞台メイクに近い化粧が出来ているという結果となった。

これより、本論文で提案した役に基づいたアイメイクデザインを生成するシステムは、インタフェースを提供することで舞台メイクに近いデザイン生成が可能だが、モチーフの特徴を反映するアルゴリズム

ムには改善の余地があることが示唆された。また、転写シールを用いることで、化粧の仕上がりを一定の水準で担保できることも分かった。

第8章 結言

本論文では、舞台演劇において手作業で行われている領域に計算機を導入するため、シンプルな操作で創作活動が可能なユーザとの対話がリアルタイムに反映されるインタフェースを提案した。リアルタイムに編集内容が反映される対話型システムを導入することで、試行錯誤がしやすくなり、アイデアの発散や新しい手法や表現の発見に繋がる。また、ミュージカルの人口の多くはアマチュアのため、複雑な設定や操作の必要がないシンプルなインタフェースの提供を目指した。その結果、専門的な知識がなくても提案したインタフェースを用いることで、一定の質が保たれた創作が行えることが分かった。

訳詞システムでは、訳詞候補を選択し、歌詞をひらがなで編集を行うだけで訳詞の作業が行えるインタフェースを提案した。また、編集した訳詞の内容はリアルタイムに譜面に反映される。ユーザが原言語や音楽に対する専門的な知識を持ち合わせていない場合でも、原曲の雰囲気を保ちつつ、歌詞の意味を反映できることを目指した。

iMake では、デザイン候補をクリックで選択し、キャンバス上で移動や回転などの編集を行うだけで役に基づいた舞台メイクのデザインが生成できるインタフェースを提案した。編集内容のキャンバス上にリアルタイムに反映される。ユーザが選択したモチーフから、アイメイクのデザインに特徴を反映し、転写シールを使うことで個人の技量による仕上がりの差をなくした舞台メイクができるようになることを目指した。

以下では本論文の各章で述べた内容を列挙し、内容を総括する。

第1章では、舞台演劇の中でも訳詞と化粧について焦点を当て、計算機を導入するために、専用のインタフェースを提供することを提案した。

第2章では、舞台演劇の創作環境としてのインタフェースを提供するにあたり、訳詞および化粧に関連するシステムの先行研究の事例について述べた。加えて、機械学習を用いた創作環境についての議論を述べた。

第3章では訳詞のアルゴリズムを考案する前段階として、4つの手法について検討をした内容を述べ、検討した手法を元にモーラ数を減らして日本語の訳詞を生成するアルゴリズムを提案した。第5章と第4章では、提案したアルゴリズムを用いて生成した日本語の訳詞を元に、ユーザが訳詞を編集するためのインタフェースの実装と評価について述べた。最終的には、システムを使用して作成された訳詞は、手作

業や GPT 3.5 による訳詞と比較して、原曲の雰囲気を保ちつつ、歌詞の意味を反映した訳詞であるという結論を得た。今後の課題として、原言語の文化的文脈をどう訳詞に反映するか、などの課題が残った。

第 6 章と第 7 章では、モチーフ画像から舞台用のアイメイクのデザインを提案するシステム “iMake” の実装と実験、評価について述べた。iMake ではモチーフの特徴を反映したアイメイクのデザインをアイシャドウ、アイライン、アイペイントの 3 つの要素に分解し、インタフェース上からそれぞれの提案されたデザインを個別に編集可能である。実験の結果、iMake を用いた化粧は舞台メイクに近いものの、生成したデザインにモチーフの特徴が反映できているかについては、化粧のデザインを生成したユーザによる評価と第三者による評価に差があったため、今後、評価差を埋めていく必要がある。

以上のように、本論文では、舞台演劇と計算機の共創を支援するために、シンプルな操作で創作活動が可能なユーザとの対話がリアルタイムに反映されるインタフェースを提案した。従来の手法と比較して、提案したインタフェースを導入した創作環境は、ユーザにとって作業がしやすい環境であることが示された。シンプルな操作により、初心者にも使いやすい環境が提供できたと言える。その一方で、より複雑な編集を行いたい、という声もあったため、今後はユーザがどの程度編集できるようにするかを模索する必要がある。また、本提案により、専門的な知識がなくとも一定の品質の創作を可能であるという結論を得られた。専用のインタフェースを導入することで、創作に必要な基本部分の候補を提供することが可能になったためだと考えられる。今後は、インタフェースに提示する候補内容の質の向上や、より使いやすいインタフェースの開発を目指したい。

計算機と共創して課題を克服することで、演劇界が発展するとともに、多くの人々が作品に触れる機会が増えるだろう。そうして、舞台がより大衆化し、身近な世界になることを願う。

謝 辞

私が研究活動を始めた2012年から今日までの12年間、大変多くの方々の御指導を頂きました。博士論文の執筆を終えるにあたり、この場をお借りして心より感謝の意を表します。

本研究を行うにあたり、終始懇切なる御指導、御鞭撻を賜りました指導教員であるお茶の水女子大学の椎尾一郎教授、伊藤貴之教授に深く感謝致します。お二人には卒業研究、修士論文に引き続き、博士論文の御指導を頂き、研究面・生活面など様々な場面で御助言を頂きました。長きにわたり、暖かく見守って頂いたことを心より感謝致します。

また、本論文の審査の過程において、お茶の水女子大学の小林一郎教授、小口正人教授、五十嵐悠紀准教授、土田修平講師には数々の御助言と御指導を賜りました。大変お忙しい中、本論文の副査をお引き受け下さいましたこと深く感謝致します。そして、NTT コミュニケーション科学基礎研究所の上村卓也様には、iMakeにおけるアイペイント作成手法の提案および実装にご協力頂きました。心より感謝致します。

社会人博士として進学するにあたり御支援下さった、Yahoo! JAPAN 株式会社、ティフォン株式会社に感謝致します。また、研究のため歌詞データの収集許可を下さった、クリプトン・フューチャー・メディア株式会社に感謝致します。

本論文の実験は、友人たち、そして、伊藤研究室の皆様にご協力して頂きました。

最後になりましたが、可愛さで私の心を支えてくれた子供たち、博士論文の執筆に理解を示し協力してくれた夫、その他御支援下さいました皆様に心より感謝致します。

A : 業績

論文誌

- [1] A. Nishimura, T. Itoh, Implementation and experiments for interactive lyrics transcreation system, *Vis. Comput. Ind. Biomed. Art* 3, 17 (2020) Jul. 2020.
- [2] 西村 綾乃, 上村 卓也, 伊藤 貴之, 椎尾 一郎, "iMake: 派手なアイメイクデザインをモチーフから作成するシステムの提案と評価, " *芸術科学会論文誌*, Vol. 22, No. 5, pp. 18:1-18:14, Dec. 2023.

国際会議

- [3] A. Nishimura, I. Siio, ConteXinger: a context-aware song generator, In Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication (UbiComp '13 Adjunct), 2013.
- [4] A. Nishimura, I. Siio, iMake: eye makeup design generator. In Proceedings of the 11th Conference on Advances in Computer Entertainment Technology (ACE '14), 2014.
- [5] A. Nishimura, I. Siio, iMake: computer-aided eye makeup. In Proceedings of the 5th Augmented Human International Conference (AH '14), 2014.
- [6] A. Nishimura, I. Siio, conteXinger: A Context-aware Song Generator to Enrich Daily Lives, In Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology (ACE2016), 2016.
- [7] A. Nishimura, T. Itoh, Implementation of an Interactive System for the Translation of Lyrics, In Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion (IUI '18 Companion), 2018.
- [8] A. Nishimura, T. Itoh, 2019. Interactive Lyric Translation System: Implementation and Experiments, In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19), 2019.

国内会議

- [9] 西村 綾乃, 椎尾 一郎, “conteXinger : 日常のコンテキストを取り込み歌う VOCALOID,” ユビキタスコンピューティングシステム研究会 第 38 回研究発表会, May 2013.
- [10] 西村 綾乃, 椎尾 一郎, “conteXinger : 日常のコンテキストを歌いあげる VOCALOID,” 音楽情報科学研究会 第 99 回研究発表会, May 2013.
- [11] 西村 綾乃, 椎尾 一郎, “conteXinger : 日常のコンテキストを取り込んだ歌声生成,” 情報処理学会 第 75 回全国大会, Mar. 2013.(学生奨励賞)
- [12] 西村 綾乃, 椎尾 一郎, “iMake : 画像からアイメイクを生成するシステムの提案,” インタラクシオン 2014, Feb. 2014.
- [13] 西村 綾乃, 氏間 可織, 安齊 クレア, 甲藤 仁美, 有本 茜, 橋本 菜摘, “Rijowarts: 技術関心を誘引するエンタテインメントシステムの開発と評価,” インタラクシオン 2015, Mar. 2015.
- [14] 西村 綾乃, 伊藤 貴之, “外国語の歌曲を訳詞するためのインターフェースの検討,” WISS 2017, Dec. 2017.
- [15] 西村 綾乃, 伊藤 貴之, “外国語の歌曲の自動訳詞システムに関する初期実装,” 音学シンポジウム 2017, Jun. 2017.
- [16] 西村 綾乃, 伊藤 貴之, “外国語の歌曲を自動訳詞するシステムのための課題と手法の検討,” JSAI 2017, May 2017.
- [17] 西村 綾乃, 伊藤 貴之, “外国語の歌曲を自動訳詞するシステムのための翻訳手法の検討,” インタラクシオン 2017, Mar. 2017.
- [18] 長谷部 礼, 西村 綾乃, “単語と画像を元にしたウェブにおけるスタイルガイド自動生成システムの提案,” WISS 2018, Sep. 2018.

その他

- [19] 西村 綾乃, 日本バーチャルリアリティ学会誌 2013 No. 4 Vol. 18, 学会参加報告, Dec. 2013.

B：参考文献

- [Yoshimura20] 吉村 佑久, ニューラル機械翻訳を用いた歌詞翻訳システム, 立命館大学修士 論文, 2020.
- [Sakurai23] 櫻井 凜, 伏見 卓恭, モーラを考慮した強化学習による歌詞翻訳, Web インテリジェンスとインタラクション研究会 予稿集, 2023, 19 巻, 第 19 回研究会, pp. 92-98, 2023.
- [Gonçalo15] H. Gonçalo Oliveira, Tra-la-Lyrics 2.0: Automatic Generation of Song Lyrics on a Semantic Domain, Journal of Artificial General Intelligence, Vol. 6 (Issue 1), pp. 87-110, 2015.
- [Gonçalo07] H. Gonçalo Oliveira, A. Cardoso, F. Pereira, Tra-la-Lyrics: An approach to generate text based on rhythm, Proceedings of the 4th International Joint Workshop on Computational Creativity, pp. 47-55, 2007.
- [Gonçalo14] H. Gonçalo Oliveira, A. Cardoso, Poetry Generation with PoeTryMe, Computational Creativity Research: Towards Creative Machines, Atlantis Thinking Machines, pp. 243-266, 2014.
- [Lerdahl85] F. Lerdahl, R. S. Jackendoff, A Generative Theory of Tonal Music, 1985.
- [Fukayama09] S. Fukayama, K. Nakatsuma, S. Sako, Y. Yonebayashi, T. Kim, S. W. Qin, T. Nakano, T. Nishimoto, S. Sagayama, Orpheus: Automatic Composition System Considering Prosody of Japanese Lyrics, Proceedings of the 8th International Conference on Entertainment Computing, pp. 309-310, 2009.
- [Marjan17] M. Ghazvininejad, X. Shi, J. Priyadarshi, K. Knight, Proceedings of ACL 2017, System Demonstrations, pp. 43-48, 2017.
- [Watanabe17] K. Watanabe, Y. Matsubayashi, K. Inui, T. Nakano, S. Fukayama, M. Goto, LyriSys: An Interactive Support System for Writing Lyrics Based on Topic Transition, Proceedings of the 22Nd International Conference on Intelligent User Interfaces, pp. 559-563, 2017.
- [Ito15] 伊藤 悠真, 寺田 努, 塚本 昌彦, Mnemonic DJ : 暗記学習のための替え歌自動生成システム, 情報処理学会論文誌, Vol. 56, pp. 2165-2176, 2015.

- [Popowich00] F. Popowich, P. McFetridge, D. Turcato, J. Toole, Machine Translation of Closed Captions, *Machine Translation*, Vol. 15, pp. 311-341, 2000.
- [Volk08] M. Volk, The Automatic Translation of Film Subtitles. A Machine Translation Success Story?, *Lang. Technol. Comput. Linguistics*, Vol. 24, pp. 115-128, 2008.
- [Ideue13] 井手上 雅迪, 対訳コーパスから生成したワードグラフによる部分的機械翻訳, 長岡技術科学大学修士論文, 2013.
- [Yamamoto13] 山本 和英, 全文を翻訳しようとする機械翻訳: ワードグラフによる部分的機械翻訳の試み (機械翻訳技術の向上), *Japio year book*, pp. 276-279, 2013.
- [Yamamoto05] 山本 和英, 池田 諭史, 大橋 一輝, 「新幹線要約」のための文末の整形, *自然言語処理*, Vol. 12, pp. 85-111, 2005.
- [Liu13] L. Liu, H. Xu, J. Xing, S. Liu, X. Zhou, S. Yan, "Wow! You Are so Beautiful Today!", *Proceedings of the 21st ACM International Conference on Multimedia*, pp. 3-12, 2013.
- [Wang06] S. Wang, Y. Wang, B. Li, Face Decorating System Based on Improved Active Shape Models, *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, pp. 65-es, 2006.
- [Tong07] W.-S. Tong, C.-K. Tang, M. S. Brown, Y.-Q. Xu, Example-Based Cosmetic Transfer, *15th Pacific Conference on Computer Graphics and Applications (PG'07)*, pp. 211-218, 2007.
- [Zhu19] F. Zhu, H. Cao, Z. Feng, Y. Zhang, W. Luo, H. Zhou, M. Song, K.-K. Ma, Semi-Supervised Eye Makeup Transfer by Swapping Learned Representation, *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 3858-3867, 2019.
- [Organisciak21] D. Organisciak, E. S. L. Ho, H. P. H. Shum, Makeup Style Transfer on Low-quality Images with Weighted Multi-scale Attention, *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 6011-6018, 2021.
- [Scherbaum11] K. Scherbaum, T. Ritschel, M. Hullin, T. Thormählen, V. Blanz, H.-P. Seidel, Computer-suggested Facial Makeup, *Comp. Graph. Forum (Proc. Eurographics 2011)*, Vol. 30, 2011.

- [Tsumura03] N. Tsumura, N. Ojima, K. Sato, M. Shiraishi, H. Shimizu, H. Nabeshima, S. Akazaki, K. Hori, Y. Miyake, Image-Based Skin Color and Texture Analysis/Synthesis by Extracting Hemoglobin and Melanin Information in the Skin, ACM SIGGRAPH 2003 Papers, pp. 770–779, 2003.
- [Huang13] C.-G. Huang, W.-C. Lin, T.-S. Huang, J. Chuang, Physically-Based Cosmetic Rendering, Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, pp. 190, 2013.
- [Rahman10] A. S. M. M. Rahman, T. T. Tran, S. A. Hossain, A. El Saddik, Augmented Rendering of Makeup Features in a Smart Interactive Mirror System for Decision Support in Cosmetic Products Selection, pp. 203-206, 2010.
- [Nakagawa11] M. Nakagawa, K. Tsukada, I. Siio, Smart Makeup System: Supporting Makeup Using Lifelog Sharing, Proceedings of the 13th International Conference on Ubiquitous Computing, pp. 483–484, 2011.
- [Kotake17] 神武 里奈, 星野 准一, 好みの顔画像の色に基づくメイクアップ支援システム, 日本感性工学会論文誌, Vol. 16, pp. 299-306, 2017.
- [Cai10] F. Cai, J. Yu, A Real-Time Interactive System for Facial Makeup of Peking Opera, Transactions on Edutainment IV, pp. 256-265, 2010.
- [Kao16] H.-L. Kao, M. Mohan, C. Schmandt, J. A. Paradiso, K. Vega, ChromoSkin: Towards Interactive Cosmetics Using Thermochromic Pigments, Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 3703–3706, 2016.
- [Lulin11] L. Ding, Digital Eyeshadow, <https://www.dlulin.com/projects/digital-eyeshadow/>, (Accessed on 2023-05-26).
- [Qu20] 瞿 佳雯, 渡辺 義浩, 高速カラープロジェクタを用いたダイナミックフェイシャルプロジェクションマッピングの試作, 研究報告コンピュータビジョンとイメージメディア (CVIM), Vol. 2020, pp. 1-4, 2020.
- [Peng23] H.-L. Peng, W. Mai, R. Sugai, J. Qu, Y. Watanabe, COLOR MACHINE, <http://www.vision.ict.e.titech.ac.jp/projects/ColorMachine/index.html>, (Accessed on 2023-05-26).

- [Vaswani17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need, arXiv, 2017.
- [Devlin18] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv, 2018.
- [Achiam23] OpenAI: J. Achiam, et al., GPT-4 Technical Report, arXiv, 2023.
- [Rombach22] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-Resolution Image Synthesis with Latent Diffusion Models, arXiv, 2022.
- [Patel03] A.D. Patel, J.R. Daniele, An empirical comparison of rhythm in language and music, *Cognition*, Vol. 87(1), pp. B35-45, 2003.
- [Sadakata03] M. Sadakata, P. Desain, H. Honing, A.D. Patel, J.R. Iversen, An analysis of rhythm in Japanese and English popular music, 2003.
- [Matsuda07] 松田 直行, 訳詞の誕生: 近代日本音楽史における訳詞家近藤朔風の位相, 駒沢短大國文, Vol. 37, pp. 15-43, 2007.
- [Port87] R.F. Port, J. Dalby, M. O'Dell, Evidence for mora timing in Japanese, *J Acoust Soc Am*, Vol. 81(5), pp. 1574-1585, 1987.
- [Ramus02] F. Ramus, Acoustic correlates of linguistic rhythm: Perspectives, In *Proceedings of Speech Prosody 2002*, Aix-en-Provence, 11-13/04/02 (pp. 115-120), 2002.
- [Matsuda13] 松田 直行, 「訳詞」とは何を訳すのか: 近藤朔風と森鷗外によるオペラ『オルフェウス』訳詞の比較研究のための序章, 駒澤日本文化 Journal Article, Vol. 7, pp. 165-192, 2013.
- [Yamamoto73] 山本 金雄, 歌唱における日本語の発音に関する一考察, 千葉大学教育学部研究紀要 第2部, Vol. 22, pp. 233-249, 1973.
- [Ikeda06] 池田 諭史, 沢井 康孝, 山本 和英, 文生成のための機能語の補完, 言語処理学会第12回年次大会発表論文集, pp. 208-211, 2006.
- [Ratcliff88] J.W. Ratcliff, D.E. Metzener, Pattern-matching-the gestalt approach, *Dr Dobbs Journal*, Vol. 13, pp. 46, 1988.

- [Abe12] 阿部 ちひろ, 伊藤 彰則, 統計的言語モデルを用いた作詞補助システムのための歌詞特徴に関する検討, 研究報告音楽情報科学 (MUS) 2012, pp. 1-6, 2012.
- [Mikolov13] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, arXiv, 2013.
- [Utiyama03] M. Utiyama, M. Takahashi, English-Japanese Translation Alignment Data, 2003.
- [Oura10] K. Oura, A. Mase, T. Yamada, S. Muto, Y. Nankaku, K. Tokuda, Recent development of the HMM-based singing voice synthesis system—Sinsy, Seventh ISCA Workshop on Speech Synthesis, 2010.
- [Wang13] Z. Wang, Study on the Importance of Cultural Context Analysis in Machine Translation, 10.1007/978-3-642-37502-6_4, 2013.
- [Hardmeier15] C. Hardmeier, On Statistical Machine Translation and Translation Theory, Proceedings of the Second Workshop on Discourse in Machine Translation, pp. 168-172, 2015.
- [Rozmyslowicz14] T. Rozmyslowicz, Machine translation: A problem for translation theory, New Voices in Translation Studies, Vol. 11, pp. 145-163, 2014.
- [watanabe18] K. Watanabe, Y. Matsubayashi, S. Fukayama, M. Goto, K. Inui, T. Nakano, A Melody-Conditioned Lyrics Language Model, In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1, pp. 163-172, 2018.
- [Dzogang10] F. Dzogang, M. Lesot, M. Rifqi, B. Bouchon-Meunier, Analysis of texts' emotional content in a multidimensional space, 2010.
- [Feng22] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, W. Wang, Language-agnostic BERT Sentence Embedding, arXiv, 2022.
- [Liang22] L. Wang, N. Yang, X. Huang, B. Jiao, L. Yang, D. Jiang, R. Majumder, F. Wei, Text Embeddings by Weakly-Supervised Contrastive Pre-training, arXiv, 2022.
- [Saegusa15] 三枝 千尋, 顔の魅力知覚に関する実験心理学的研究, University of Tokyo(東京大学), <https://ci.nii.ac.jp/naid/500001043809> 2015.

- [Morikawa13] 森川 和則, 錯視としての化粧効果の測定と考察 (特集 見た目の科学とアンチエイジング), フレグランスジャーナル : 化粧品科学研究開発専門誌, Vol. 41, pp. 55-61, 2013.
- [Baudouin04] J.-Y. Baudouin, G. Tiberghien, Symmetry, averageness, and feature size in the facial attractiveness of women, *Acta Psychologica*, Vol. 117, pp. 313-332, 2004.
- [Morikawa15] 森川 和則, 2-1 化粧による顔の心理効果 ～顔錯視研究の観点から～, 映像情報メディア学会誌, Vol. 69, pp. 842-847, 2015.
- [Muto19] H. Muto, M. Ide, A. Tomita, K. Morikawa, Viewpoint invariance of eye size illusion caused by eyeshadow, *Frontiers in Psychology*, Vol. 10, pp. 1510, 2019.
- [Wang11] H. Wang, H. Zhang, N. Ray, Clump splitting via bottleneck detection, *Proceedings - International Conference on Image Processing*, pp. 61-64, 2011.
- [Gatys16] L. A. Gatys, A. S. Ecker, M. Bethge, Image Style Transfer Using Convolutional Neural Networks, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2414-2423, 2016.
- [Deng09] J. Deng, W. Dong, R. Socher, L. Li-Jia, K. Li, F. Li-Fei, ImageNet: A large-scale hierarchical image database, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 248-255, 2009.
- [Mikolov18] T. Mikolov, E. Grave, P. Bojanowski, C. Puhersch, A. Joulin, Advances in Pre-Training Distributed Word Representations, *Proceedings of the International Conference on Language Resources and Evaluation*, 2018.
- [Wang22] H. Wang, H. Xie, K. Miyata, ILoveEye: Eyeliner Makeup Guidance System with Eye Shape Features, *arXiv*, 2022.

Advanced Science,
Graduate School of Humanities and Sciences,
Ochanomizu University,
Ayano Nishimura