歌詞が持つ言葉の印象に合致した自動文字 PV 生成方式

前井涼花† 岡田龍太郎† 中西崇文†

†武蔵野大学データサイエンス学部 〒135-8181 東京都江東区有明 3 丁目 3-3

E-mail: †s1922071@stu.musashino-u.ac.jp, {ryotaro.okada, takafumi.nakanishi}@ds.musashino-u.ac.jp

あらまし 本稿では，歌詞が持つ言葉の印象に合致した自動文字 PV 生成方式について示す．本方式は，入力された歌詞から，各歌詞のフレーズとその印象を表す印象語との類似度を計算し、その歌詞のフレーズの印象を推定し，その印象に合致したフォントの種類，映像エフェクトを決定し，文字 PV として自動生成するものである．本方式は，歌詞の印象に合致したフォントの種類，映像エフェクトを決定することにより，楽曲の印象に合致した文字 PV を自動的に作成することが可能となる．本方式は，自作の楽曲の発信者にとって，その楽曲に合致した動画の自動生成を容易にすること可能にし，自作の楽曲を映像付きのリッチなコンテンツとして配信する一助となりうる．

キーワード 文字 PV，動画自動生成，楽曲，歌詞

1. はじめに

近年，DTM ソフトウエア，関連ハードウエア，音声合成技術の廉価化，発達により，個人でも高品質な楽曲製作を行うことが可能となりつつある．これらの製作された楽曲はインタネット上で公開され，膨大な量のオリジナル楽曲が存在する．インタネット上の膨大な量のオリジナル楽曲を対象として，ユーザがこれらのオリジナル楽曲を対象として，単なる聴取に付加価値をつけた新たな楽しむ方法の実現が重要となってきている．

インタネット上では，楽曲とともに歌詞を表現した文字 PV(Promotion Video)コンテンツが散在するようになった．文字 PV とは，画面上で歌詞となる言葉を動作させたり様々なフォント，エフェクトをかけたりすることによって，楽曲のイメージにあった文字からなるビデオコンテンツを指す．楽曲の進行に合わせて，音声，もしくは歌詞の言葉が持つ印象に合致したフォント，エフェクトを用いて歌詞の文字列を表現することにより，楽曲をより深く楽しむためのコンテンツとなっている．文字 PV は，文字と単純な図形のみから構成されていることが多く，動画コンテンツとして，要素がシンプルであるため，どの言葉にどのようなフォントで表示し，どのようなエフェクトをかけるかは，文字 PV を製作するクリエイタの専門知識に依存する．一方で，動画製作の専門知識を持たない楽曲を製作するクリエイタにとっては，自身が製作する楽曲に合致した文字 PV を気軽に創ることで，自身の楽曲に付加価値を付けて発信したいという要望が高まっている．任意の楽曲に対して，その楽曲や歌詞の印象に合致した文字 PV の自動生成が求められている．

本稿では，歌詞が持つ言葉の印象に合致した自動文字 PV 生成方式について示す．本方式は，入力された歌詞から，各歌詞のフレーズとその印象を表す印象語との類似度を計算し、その歌詞のフレーズの印象を推定し，その印象に合致したフォントの種類，映像エフェクトを決定し，文字 PV として自動生成するものである．本方式は，歌詞の印象に合致したフォントの種類，映像エフェクトを決定することにより，楽曲の印象に合致した文字 PV を自動的に作成することが可能となる．

本方式は，自作の楽曲の発信者にとって，その楽曲に合致した動画の自動生成を容易にすること可能にし，自作の楽曲を映像付きのリッチなコンテンツとして配信する一助となりうる．本方式を実現することにより，インタネット上の楽曲に自動的に効果的な動画を生成することが可能となり，そのような動画コンテンツが増大することにより，単なる聴取に付加価値をつけたユーザ体験を提供することが可能になると考えられる．

本稿は，次のように構成される．2 節では，関連研究について示す．3 節では，本方式である歌詞が持つ言葉の印象に合致した自動文字 PV 生成方式について示す．4 節では，本方式を実現する実験システムを構築し，実験結果を示す．5 節で，本稿をまとめる．

2. 関連研究

音楽動画における自動文字挿入に関する研究について，野中ら[1]は，プロモーションビデオなどの音楽動画の印象に合わせて既存のフォントを組み合わせ新たなフォントを生成し，さらにそのフォントでその音楽動画に歌詞を自動挿入している．本研究では，プロモーションビデオなどの音楽動画が存在しない楽曲を対象として，楽曲の歌詞の印象に基づき，適した既存のフォントを選択し，選択されたフォントで構成される文字主体とした歌詞を表現する動画（文字 PV）を新たに生成している．

梅村[2]らは，用意された画像について，あらかじめ

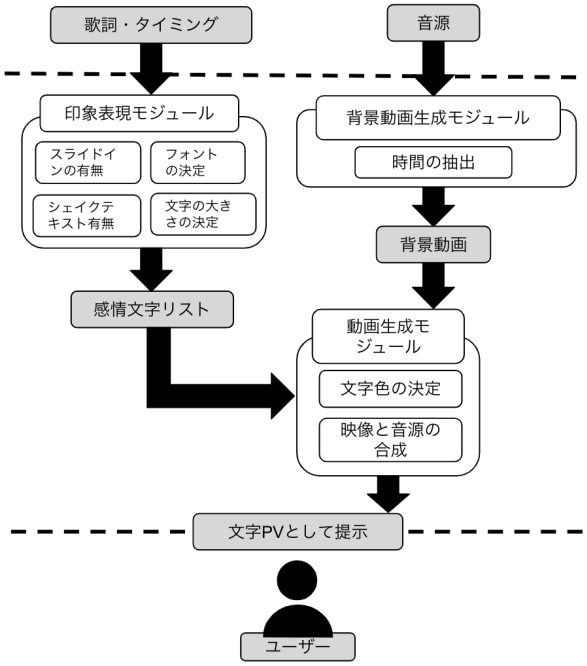

図 1　全体像

人手で印象を表すラベルを付与した上で画像と歌詞の類似度を計算することで，その類似度の高い画像群を利用した楽曲に合致したスライドショーを生成するシステムを実現している．本研究では，歌詞とフォントの印象の類似度を計算し，歌詞の文字情報から構成される音楽動画を自動生成している．

文字の印象に関する研究について，石井ら[3]はテキストの印象によってフォント・色・大きさが自動決定されるシステムを提案している．石井らは，任意の段落ごとに TF-IDF によって抽出された上位単語のうち感性特徴量が定義されている単語 15 語を選出し，ベクトル化し，フォント・文字色のベクトルとの内積を計算することで，上位のフォントを採用し自動装飾するシステムである．本研究では，石井ら[3]の研究を発展させ,類似度によりフォントを選択するだけでなく，印象の類似度により，動画のエフェクトを決定し，文字 PV の自動生成を可能としている．

3. 自動文字 PV 生成方式

3.1. 全体像

本節では本研究の全体像について述べる．全体像を図 1 に示す．

本方式は，印象表現モジュール，背景動画生成モジュール，動画合成モジュールからなる．本方式は，入力データとして，楽曲音声データ，歌詞テキストデータ，歌詞のそれぞれの部分を表示させるタイミングを示す時間を記述したメタデータの 3 つを与え，出力として，歌詞の印象に合致した文字 PV を生成する．印象表現モジュールは，入力データのうち，歌詞とその出現するタイミングから,感情文字リストを出力する．感情文字リストとは，歌詞の文字列に対して，文字の大きさ，フォント，スライドインの有無，シェイクテキストの有無の属性を付加したものである．背景動画生成モジュールは，音源と入力とし，音源の長さに対応する背景動画を出力する．動画合成モジュールは，感情文字リストと背景動画を入力とし，文字 PV を出力する．

表 1: 印象クラスとフォントの関係

印象クラス	フォント
堂々	装甲明朝
元気が出る	ニコ角
切ない	ほっこり
激しい	851 チカラヅヨク
滑稽	青柳疎石フォント
かわいい	JK ゴシック M

3.2. 印象表現モジュール

本節では，印象表現モジュールの詳細について述べる．印象表現モジュールは，入力された歌詞の文字列に対して，視覚上の表現形態を決定するモジュールである．本モジュールは，文字の大きさの決定，フォントの決定，スライドインの有無の決定，シェイクテキストの有無の決定の 4 つのサブモジュールから構成される．

3.2.1. フォントの決定

フォントは任意の 6 種類を決定しそれぞれ印象クラスに対応付けている．印象クラスは大野ら[4]が使用している音楽ワークショップである MIREX で用いられている 5 つの印象クラスを使用する．大野らはニコニコ動画では「かわいい」と感じる楽曲やタグが多く存在することから, MIREX の 5 つの印象クラスに加え「かわいい」を追加している．本研究は文字を主体とした歌詞を表現する動画の生成するものであり，ニコニコ動画に多く投稿されているボーカロイド楽曲の動画生成に有効であるとされる．そのためニコニコ動画の印象クラスに対応させる．6 種類の印象クラスに対応するフォントを表 1 に示す．フォントは印象に合致する範囲であれば任意に変更可能である．

フォントは 3.2 節の印象表現モデルで述べたように任意の 6 種類であり印象クラスに対応している．入力された歌詞と印象クラスの類似度を, Word2Vec を用いて計量する．Word2Vec の学習済みモデルには，鈴木ら[5]の提供する，Wikipedia の記事から生成されたモデルを用いる．Word2Vec モデルを用いて，楽曲の歌詞をユーザが入力した単位ごとにベクトル化する．また，印象クラスの語もそれぞれ同様にベクトル化する．な

表 2 文字の大きさの決定の条件

類似度	文字の大きさ
0.5 よりも大きい	100
0.5 以下	50

表 3 スライドインの判定条件

一秒あたりの文字数	表示する時間	スライドイン
4 文字より多い	1 秒より長い	あり
	1 秒以下	なし
4 文字以下	1 秒より長い	なし
	1 秒以下	

表４　シェイクテキストの決定の条件

類似度	シェイクテキスト	位置
0．3 より大きい	あり	変化あり
0．3 以下	なし	変化なし

お，印象クラスのうち「元気が出る」は，他の語と違い二文節からなるため，統一のため「元気」として扱うこととした．入力した歌詞と印象クラスの各語との類似度を，これらのベクトルのコサイン類似度として計量し，最大の類似度を持つ印象クラスに対応するフォントを，その歌詞のフォントとして採用する．

3.2.2. 文字の大きさの決定

音源のタイトルと類似度が高い単語は印象の表現において重要であると仮定し，類似度に応じて文字の大きさを変化させる．文字の大きさは 50pt と 100pt の 2 種類とし，曲名と類似度の高い単語を表示させる場合には文字の大きさを 50pt とする．類似度の計量方法はフォントの種類選択と同様である．表２のように類似度が 0.5 よりも大きい場合文字の大きさを 50pt とする．この閾値は任意に設定することが可能である．

3.2.3. スライドインの有無の決定

一度に表示させる歌詞が多い場合，その歌詞は短い時間に文字量に詰め込まれていることを意味し，言葉のテンポが早いことを意味する．そのため，文字量が多い場合には，そのスピード感を表すために，文字をスライドインさせる．本実装では，スライドインの表現として，文字を左から右に流して表示している．

入力した歌詞がスライドインになる条件は，一秒あたりの文字数が 4 文字より多く，かつ歌詞を表示させる時間が一秒よりも長い場合である．スライドインの判定条件を表 3 に示す．

3.2.4. シェイクテキスト

シェイクテキストは文字が揺れるエフェクトを指す．シェイクテキストは“れっくぷらす/rec plus”の動画[6]によれば，バラエティ番組において危機感を表す場合によく使用されると述べられている．そのため，本方式では，危機感を感じさせる語を含む歌詞に対して，シェイクテキストを適用する．具体的には，危機感という単語と歌詞の類似度が高い時に実装される．類似度の計算はフォントの種類・文字の大きさと同様に求める．表 4 に示す通り，類似度が 0.3 より大きい時，そのときの再生時間によって位置が変化し揺れを生成している．本実装では歌詞をシェイクテキストとして，文字列を横方向に揺らすこととした．

3.3. 背景動画生成モジュール

本節では背景動画生成モジュールについて述べる．背景動画は入力された音源の長さに合わせて自動生成される．背景は白黒を基調としており，その中で変化を付ける．本実装では，周期的に中心にある正方形が大きくなる動きを持たせることとした．正方形は白と黒が交互に現れる．

3.4. 動画合成モジュール

本節では感情文字リストと背景動画の合成について述べる．文字の色は白か黒とし，背景が白である場合は黒，背景が黒である場合は白となる．文字同士が重なっている場合は白と黒が重なっている部分のみ反転する．背景と文字を合成した後，音源を合成し文字 P V が出力される．

4. 実験

本節は本方式の評価実験について述べる．

実験 1 では提案方式による文字 PV 生成の実現例を示す．実験２では提案方式で生成された文字 P V と単調な文字 P V の比較実験を行い，生成された文字 P V の印象評価を行うことで印象表現が適切に行えていることを示す．

4.1. 実験環境

3 節で提案したシステムを実装し，入力する楽曲として，「拝啓ドッペルゲンガー」(作詞作曲：kemu feat. GUMI)と，「蜃気楼に求め」(作詞：ウォルピスカーター，作曲：SILVANA)の 2 曲を用いて文字 P V を生成した．「拝啓ドッペルゲンガー」はテンポが速く歌詞も多い楽曲であり，「蜃気楼に求め」はテンポが遅いバラードである．

4.2. 実験 1: 文字 PV 生成の検証

印象表現モジュールを実装し，フォントの変化や文字の大きさ，文字の動きの確認を行なった．生成された動画のある時点でのスクリーンショットを図 2,3 に

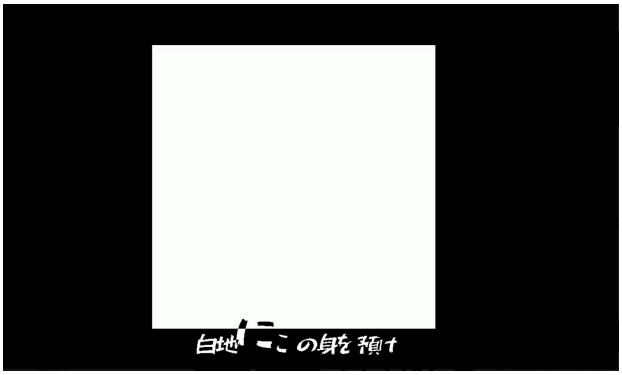

図 2　実験 1 の結果

図 3　実験 1 の結果

表 5　アンケート項目

Q1	歌詞の意味と選択されているフォントやエフェクトが合致しているか
Q2	PV としての完成度はどうか
Q3	その他改善点

示す.

図からわかるようにきちんと歌詞によってフォントの変化や文字の大きさの変化を見ることができる. 背景の色や文字の重なりによって変化する文字色の決定も適切に行われている. シェイクテキスト・スライドインに関しても実装できていることが確認できた.

4.3. 実験 2:印象表現の検証

生成した文字 PV が歌詞の持つ印象に合致しているか評価するための実験を行った. 歌詞の持つ印象に対応していない単調な文字 PV と本システムを使用して生成した文字 PV の比較実験を行った. 被験者は, 単調な文字 PV と本方式で生成された文字 PV のそれぞれ 2 本を視聴した上で表 5 の三項目について回答した. Q1・Q2 は, 1 ～ 5 の五段階評価である. 1 に近いほど本システムの出力した PV の方が良いと感じ, 5 に近いほど単調な文字 PV の方が良いと感じた評価とする. Q3 は本方式の改善点を探るため自由記述で回答してもらった.

実験には 10 代から 20 代の男女 19 人が参加した.

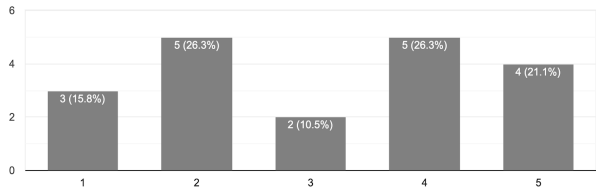

図 4 「昼気楼に求め」における歌詞の意味とフォントやエフェクトの合致

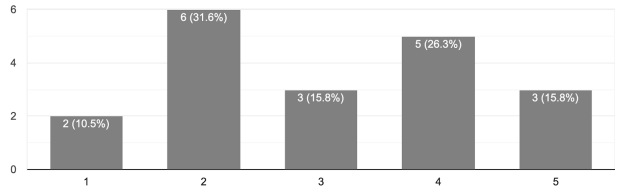

図 5「昼気楼に求め」における PV の完成度

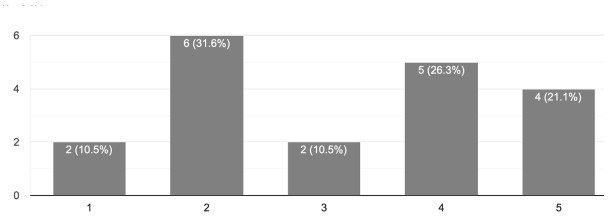

図 6「拝啓ドッペルゲンガー」における歌詞の意味とフォントやエフェクトの合致

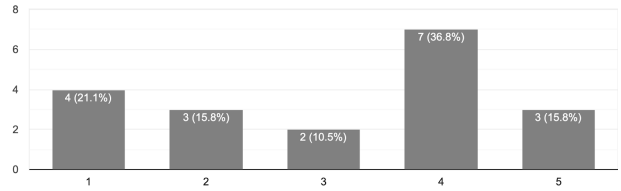

図 7「拝啓ドッペルゲンガー」における PV の完成度

実験結果を図 4,5,6,7 に示す.

図 4,5 は「昼気楼に求め」の音源に対して生成された動画に関するアンケート結果である. 図 6,7 は「拝啓ドッペルゲンガー」の音源に対して生成された動画に関するアンケート結果である. 図 4 と図 6 から, 歌詞の意味と選択されているフォントやエフェクトが合致しているかという点については, 両楽曲とも, 本システムによって作られた文字 PV と, 単調な文字 PV と

図８　文字が見えづらい例

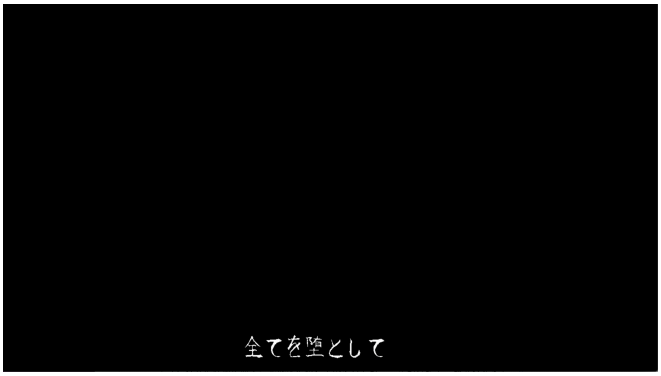

図９　フォント変更後

で，評価が別れていることが分かる．また，図 5 と図 7 により，PV の完成度についても，両楽曲とも評価が別れていることが分かる．図 7 においては，「拝啓ドッペルゲンガー」を入力とした場合に，やや単調な文字 PV の方が評価が高く，完成度において本システムによる文字 PV より単調な文字 PV の方が高評価となっている．

Q３の回答としては，高評価の意見として，「見ていて気持ちが良かった」，「歌詞とエフェクトが合致しているように見えた」，「ドッペルゲンガーの所が強調されていていいと感じた」といったものがあった．低評価の意見として，「フラットな雰囲気のワードもくせのあるフォントでごちゃごちゃした印象を受けた」，「文字が見辛い時がある」，「歌詞とエフェクトがあまり合っていないことやエフェクトの規則性がわからないことが気になってしまい、歌に集中できなかった。」といったものがあった．特に，本システムによって生成した，印象表現を加えた PV に対して，「文字が見辛い」というコメントが複数あった．

4.4. 考察

実験 1 により，楽曲および歌詞とそれを表示させるタイミングを入力することにより，文字 PV を生成する例を示すことが出来た．

実験 2 により，歌詞の印象表現の検証を行ったところ，高評価である意見が複数あったことから，本システムは歌詞の印象を文字ＰＶに反映させることができたと言える．一方で，低評価の意見もあることから，PV の完成度において改善の必要があることが示された．

PV の完成度における評価が低くなった主な原因としてフォントの選択が良くなかったことが挙げられる．細字のフォントが選択された際，線が消えてしまい歌詞が読みづらくなっている．文字が見づらくなっている状況の例を図 8 に示す．歌詞が読みづらいことで完成度が低く統一性がない印象を受け，単調な動画の印象評価が高くなっているのではないかと予想する．これは選択するフォントを変更することで改善が可能と考え,フォントを変更したものを図 9 に示す．

5. おわりに

本稿では，歌詞が持つ言葉の印象に合致した自動文字 PV 生成方式について述べた．本方式は，入力された歌詞から，各歌詞のフレーズとその印象を表す印象語との類似度を計算し、その歌詞のフレーズの印象を推定し，その印象に合致したフォントの種類，映像エフェクトを決定し，文字 PV として自動生成するものである．本方式は，歌詞の印象に合致したフォントの種類，映像エフェクトを決定することにより，楽曲の印象に合致した文字 PV を自動的に作成することが可能となる．

本方式は，自作の楽曲の発信者にとって，その楽曲に合致した動画の自動生成を容易にすることを可能にし，自作の楽曲を映像付きのリッチなコンテンツとして配信する一助となりうる．

また，本稿では，本方式を実現する実験システムを構築し，自動生成された文字 PV についてアンケート調査を行うことで有効性の検証を行った．

今後の課題としては，さらなる映像エフェクトの導入手法の実現，動画製作の現場のクリエイタが持つ専門知識の知識ベース化とそれを用いた自動文字 PV 生成方式への発展，本方式を Web アプリへ展開することによる文字PVポータルサービスの実現が挙げられる．

参　考　文　献

[1] 野中滉介,斉藤絢基,中村聡史,“音楽印象と同期した歌詞フォント融合による印象強調手法”，情報処理学会研究報告，IPSJ SIG Technical Report，Vol.2018-EC-50，No.35，2018/12/22

[2] 梅村允康,保利武志,嵯峨山茂樹,“Word2Vec を用いて歌詞と写真を対応づけたスライドショー生成システム”，情報処理学会第 81 回全国大会

[3] 石井千賀,倉林修一,清木康，“文書印象表現単語とフォントの感性的相関量計量による動的文書装飾システム”，DEIM Forum 2011 E8-3/ Chika ISHII,Shuichi KURABAYASHI,Yasushi KIYOKI, A Dynamic Text-Decoration System for Text by

Calculating Correlation between Impressive-Words and Fonts

[4] 大野直紀,土屋駿貴,中村聡史,山本岳洋, “独立した音楽と映像に対する印象評価と音楽動画の印象の関係性に関する研究”, 情報処理学会論文誌,Vol.59,No.3,929–940,(Mar. 2018)

[5] 鈴木正敏, 松田耕史, 関根聡, 岡崎直観, 乾健太郎. Wikipedia 記事に対する拡張固有表現ラベルの多重付与. 言語処理学会第 22 回年次大会 (NLP2016), March2016. http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/

[6] れっくぷらす/Rec Plus, シェイク テキストエフェクト 文字を揺らす方法【Premiere Pro / プレミアプロ チュートリアル】