ユーザの印象を反映する対話型演奏表情生成†

鈴木 未央*1・鬼沢 武久*2

本論文では、ユーザとの対話により言葉で表された印象を反映する演奏表情生成システムMUSAI（MUSical expression generation system by Adjective with Interaction）を提案する。MUSAIは、演奏表情と印象の関係を表す感性空間、ファジィルールを持ち、入力された印象から感性空間上での座標値を求め、ファジィ推論によって演奏表情のパラメータを決定し、演奏表情を生成する。生成された演奏表情に満足できない場合、ユーザは副詞と形容詞・形容動詞からなる修正指示語を入力し、感性空間上の座標値を変えることで演奏表情を修正する。有効性を評価するために、MUSAIを用いて実験参加者が演奏表情を生成し、評価する実験を行う。実験の結果、MUSAIにより、各実験参加者の印象を十分に反映した演奏表情を生成できること、実験参加者が演奏表情が次第に修正されていると知覚していたことがわかった。

キーワード：演奏表情、印象、感性空間、ファジィ推論、インタラクション

1. はじめに

演奏表情とは、演奏によってある印象や感情を表現するために、楽譜に記述されている速度記号や強弱記号といった指示記号から逸脱した演奏のことを指す。同じ楽曲でも演奏されるごとに、また、演奏者によって印象が異なって聞こえるのは、演奏時の状況や演奏家ごとの表現の違いが影響し、演奏表情が変化しているためである。したがって、楽譜通りの演奏が必ずしも表情豊かな演奏とは限らない[1]。さて、楽曲の演奏を考えたとき、どのような演奏形態であっても必ず演奏者が存在し、オーケストラや吹奏楽といった演奏形態の場合には演奏者以外に演奏に関わる立場として指揮者が存在する。また、特に、アマチュアの演奏団体では、指揮者が指導者を兼ねており、合奏練習や個人練習の際に演奏指導を行うこともある。指揮者、指導者（以降、まとめて指導者とする）は、練習の際に、楽曲にどのような表情を付けて演奏するかを演奏者とやり取りすることで、目標とする演奏を作り上げる。具体的には、まず、指導者が演奏する楽曲にどのような表情を付けたいかを演奏者へと言葉で伝え、目標となる演奏表情を演奏者と共有する。演奏者は指導者から伝えられた目標に沿うように演奏を行う。このとき、演奏者の演奏が、指導者が思い描いていたもの

と異なる場合には、指導者は演奏者に対してさらなる指示を行う。本論文では、このような指導者と演奏者のやり取りのように、言葉を用いてユーザと対話することで演奏表情を生成、修正するシステム Musical Expression Generation System by Adjective with Interaction (MUSAI)を考える。ユーザが指導者、MUSAIが演奏者と見立て、ユーザが楽曲に付けたい演奏表情の印象を形容詞または形容動詞で入力すると、MUSAIはその印象を反映する演奏表情を生成する。そして、MUSAIが生成した演奏表情に印象が反映されていないとユーザが感じる場合には、ユーザはMUSAIに対して修正の指示を言葉で行う。本論文では、印象と演奏表情の関係を感性空間とファジィルールで表現し、入力された印象を表す言葉の感性空間上での座標値を求め、座標値から演奏表情を構成する要素のパラメータ値をファジィ推論により決定することで演奏表情を生成する。

コンピュータで演奏表情を生成する研究はこれまで、ルールベース、事例ベース、学習モデルベースといった様々なアプローチで行われており、事例となる演奏が持つルールの抽出や特徴を再現するものと、ユーザの意図を取り入れてユーザに演奏生成の新しい体験を提供するものがある[2]。従来研究のうち、ルールベース、学習モデルベースの研究では、適用するルールをユーザが選択する、または、ルールの重み付けを行うことで、ユーザの意図を取り入れた演奏表情を生成している[2]。しかし、ルール数が多くなるとつれ、意図を反映させるためのルール選択や重み付けにおいてユーザの負担が大きくなると考えられる。ユーザの意図を取り入れる方法として、演奏表情に反

† Musical Expression Generation Reflecting User's Impression by Interaction

Mio SUZUKI and Takehisa ONISAWA

*1 釧路工業高等専門学校

National Institute of Technology, Kushiro College

*2 筑波大学

University of Tsukuba

映させたい感情や印象、雰囲気を言葉で表して入力することで、ユーザの意図を取り入れる研究もある[2, 3]。言葉を用いることでユーザの意図を演奏表情に反映させる研究[4, 5, 6]では、言葉と演奏表情を関連付けているが、いずれの研究も扱う言葉はあらかじめ用意されたものであるため、ユーザの意図は用意された言葉の範囲でしか反映させることができず、用意されていない言葉で表される意図への対応が難しい。他にも、ユーザの意図を指揮棒の動きで表し、演奏表情に反映させる研究がある[7]。ユーザの意図を実際の指揮者が行うように指揮棒の動きで表すことで演奏のテンポと音量へ反映し、さらに Canazza らの演奏表情生成システム[8]を用いてユーザの意図を音色に反映させている。しかし、指揮者がいる演奏形態での演奏経験がなければ、指揮棒の動きが表す意図を理解するのは難しいことから、指揮棒の動きからどのような意図で演奏表情を生成したのかを説明することは難しい。本論文で述べる MUSAI は、演奏表情に反映させたい印象を形容詞または形容動詞であれば自由に入力できるものとし、さらに、演奏表情の修正においても言葉を用いることで、ユーザがどのように演奏表情を修正したいかが感性空間上で説明可能となる。また、本論文と同様に、ソフトコンピューティング手法を用いて個人の感性を反映させる研究として、秋口の自動楽曲生成の研究がある[9]。秋口の研究では、ニューラルネットワークを用いて個人の感性をモデル化し、入力された印象を反映するような楽曲を遺伝的アルゴリズムによって探索、生成している。しかし、ニューラルネットワークによる感性のモデル化では、楽曲と印象の具体的な関係を知ることが難しい。本論文では、ユーザが入力した印象を表す言葉を演奏表情に反映するために、感性空間とファジィルールをあらかじめ用意し、感性空間上の座標値と演奏表情を表すパラメータの値の関係を説明可能な形で表す。

本論文の構成は以下の通りである。2章では言葉で表されたユーザの印象を反映する演奏表情生成システム MUSAI について述べる。3章では MUSAI の有効性を検証するための実験とその結果について述べる。最後に、4章において本論文の結論を述べる。

2. 演奏表情生成の概要

図1に本論文で述べる演奏表情生成システム MUSAI の概要を示す。ユーザは演奏表情を付与したい楽曲フレーズと、演奏表情で表現したい印象を「激しい」「元気な」などの形容詞または形容動詞で入力する。楽曲フレーズは、楽譜表記フォーマットである MusicXML [10]形式のデータで入力され、印象を表現する形容

図1 演奏表情生成の概要

詞、形容動詞を本論文ではイメージ語と呼ぶ。実際の演奏を考えたとき、演奏表情によって表される印象の言語表現は多様であることから、ユーザが入力するイメージ語は形容詞、形容動詞であれば自由に入力することとする。イメージ語の感性空間（以後、単に感性空間と呼ぶ）上の座標値をイメージ推定処理で求め、座標値から演奏表情パラメータをファジィ推論により求める。求められた演奏表情パラメータと楽曲フレーズから演奏表情のMIDIデータを作成し、ユーザに提示する。提示された演奏表情がユーザにとって満足できないものであった場合、ユーザは「もっと激しい」「少し元気でない」のような副詞とイメージ語を組み合わせた修正指示語を入力する。入力された修正指示語から、修正後の候補となる演奏表情の座標値が複数求められる。求められた座標値から再度ファジィ推論を行い、演奏表情を生成してユーザに提示する。ユーザは提示された修正後の候補となる演奏表情について、イメージ語が反映されているかどうかの評価を行う。演奏表情の修正と評価を繰り返すことで、各ユーザの印象を十分に反映した演奏表情を生成する。対話的に演奏表情を修正することで、実際の演奏における演奏表情と同様に、同じユーザでもそのときの状況に合わせて、同じ印象でも異なる演奏表情を生成することができるようになる。

2.1 演奏表情のパラメータ

本論文では演奏表情を構成する要素をテンポ、音量、音の長さの三つとする。これらは、楽曲の調や旋律といった複数の要素の中で、演奏時に演奏者が変化させることができる要素である。要素はそれぞれ表1に示すパラメータで表現する。楽曲フレーズの始めから終わりまでの音符数を n としたとき、要素はそれ

表1 演奏表情のパラメータ

要素	パラメータ	値	意味
テンポ	<i>TempoBase</i>	[40, 208]	演奏全体の基準となるテンポ
	<i>TempoRange</i>	[0.0, 0.6]	テンポの変化幅
	<i>TempoVar</i>	<i>constant, decrease, increase</i>	テンポの変化の仕方を示すフラグ
音量	<i>VelocityBase</i>	[16, 127]	演奏全体の基準となる音量
	<i>VelocityRange</i>	[0.0, 0.7]	音量の変化幅
	<i>VelocityVar</i>	<i>constant, decrease, increase</i>	音量の変化の仕方を示すフラグ
音の長さ	<i>LengthSign</i>	-1, 0, 1	音を長く演奏するか短く演奏するかを示す
	<i>LengthBase</i>	[1.5, 10.0]	楽譜からどれだけ短く・長くなるかを示す
	<i>LengthRange</i>	0	音の長さの変化幅
	<i>LengthVar</i>	<i>constant</i>	音の長さの変化の仕方を示すフラグ

ぞれベクトル $\vec{T} = (t_1, t_2, \dots, t_n)$, $\vec{V} = (v_1, v_2, \dots, v_n)$, $\vec{L} = (l_1, l_2, \dots, l_n)$ で表現される. 各ベクトルの成分である t_i, v_i, l_i は楽曲フレーズ中の音符一つのテンポ, 音量, 音の長さを示す.

2.2 感性空間の構築

印象と演奏表情の関係を表す感性空間を構築するために, 予備実験を行う.

2.2.1 空間構築のための予備実験

予備実験では, パラメータをランダムに設定して演奏表情を自動生成し, 実験参加者に対して生成された演奏表情をランダムに提示する. 参加者は提示された演奏表情を聞き, その演奏表情から受ける印象について, 先行研究[11]を参考に選出した表2に示す43個

の形容詞対を用いて, 5段階SD(Semantic Differential)法[12]で評価する. 形容詞対の評価尺度の例を図2に示す. 評価尺度上のラベルは, 文献[13]を参考に, 中央から両極に向かって「どちらでもない」「どちらかというと思う」「そう思う」と設定する. 演奏表情を付与する楽曲はベートーヴェンのピアノソナタ第14番Op. 27 第2楽章で, 楽曲の冒頭から15小節目までとする. 演奏はSMF(Standard MIDI File)形式のデータで作成し, MIDI音源の音色はAcoustic Grand Pianoに設定する. 実験参加者は19歳から25歳の男女16名である.

図2 予備実験で用いる評価尺度の例

表2 予備実験で用いた形容詞対

かわいい	—	憎らしい	子供らしい	—	大人っぽい
ほがらかな	—	陰気な	気楽な	—	気がかりな
楽しい	—	苦しい	愉快的な	—	不愉快的な
はつらつとした	—	病弱な	華やかな	—	地味な
にぎやかな	—	寂しい	活動的な	—	内向的な
大胆な	—	臆病な	情熱的な	—	冷静な
激しい	—	穏やかな	ゴージャズな	—	簡素な
つやつぱい	—	色気のない	贅沢な	—	質素な
力強い	—	弱弱しい	ワイルドな	—	おとなしい
アンティークな	—	新しい	重厚な	—	軽薄な
神聖な	—	俗っぽい	精密な	—	アバウトな
都会的な	—	田舎的な	シックな	—	野暮な
素直な	—	意固地な	おおらかな	—	こせこせした
ナチュラルな	—	人工的な	女性的な	—	男性的な
優雅な	—	粗野な	上品な	—	下品な
洗練された	—	ださい	わかわかしい	—	年老いた
スピーディーな	—	ゆったりした	モダンな	—	クラシックな
すっきりした	—	憂鬱な	健康的な	—	病弱な
繊細な	—	図太い	平和な	—	騒々しい
淡白な	—	濃厚な	なごやかな	—	陰しい
なじみやすい	—	なじみにくい	ロマンティックな	—	リアリスティックな
すがすがしい	—	うっとうしい			

2.2.2 感性空間を構築する因子

予備実験で得られた延べ177個の演奏表情に対する評価結果について、形容詞対の尺度を変量として因子分析を行い、演奏表情の感性空間を構築する。因子の抽出には最尤法を用いる。初期の固有値および回転後の固有値からスクリー基準で因子数を判断した場合、抽出する因子数は3となるが、第3因子に含まれる形容詞対は1つとなり、因子として扱うことができない。また、因子数を4とした場合、第4因子の因子負荷量が第1因子から第3因子の因子負荷量と比較すると値が小さく、第4因子の固有値も小さいという結果となる。そのため、本論文では、抽出する因子数を2と判断し、バリマックス回転を用いて各形容詞対の因子負荷量を求める。各形容詞対の中で共通性の値が0.16以下である「色気のない-つやばい」「濃厚な-淡白な」「大人っぽい-子供らしい」「軽薄な-重厚な」を分析から除外し、因子負荷量の絶対値が0.7以上の値を持つ形容詞対13個が、2因子を抽出した場合に因子に含まれることとなる。表3に因子分析で抽出さ

表3 抽出された2因子の因子負荷量

形容詞対	因子	
	第1因子	第2因子
病弱な-はつらつとした	0.776	0.139
寂しい-にぎやかな	0.810	-0.037
臆病な-大胆な	0.723	-0.090
穏やかな-激しい	0.735	-0.252
弱弱い-力強い	0.700	-0.013
ゆったりした-スピーディーな	0.723	-0.126
地味な-華やかな	0.730	0.241
内向的な-活動的な	0.871	-0.014
おとなしい-ワイルドな	0.760	-0.273
粗野な-優雅な	-0.081	0.723
なじみにくい-なじみやすい	0.393	0.706
うっとうしい-すがすがしい	0.316	0.709
下品な-上品な	-0.172	0.770
固有値	10.156	8.299
寄与率 (%)	26.041	21.280
累積寄与率 (%)	26.041	47.321

図3 活動性因子軸

図4 気品因子軸

れた2因子の寄与率と因子に含まれる形容詞対を示す。第1因子は動きを連想させる形容詞対で構成されていることから「活動性因子」とし、第2因子は「粗野な-優雅な」「下品な-上品な」といった品を感じさせる形容詞対で構成されていることから「気品因子」とする。本論文では、活動性因子と気品因子の各軸を $[-1.0, +1.0]$ の範囲で図3、図4のように改めて定義し、二つの因子を軸とした2次元空間を演奏表情の生成に用いる感性空間とする。

2.3 イメージ推定処理

感性空間上でのイメージ語の座標値を求めるためにイメージ推定処理を行う。イメージ推定処理では共立共起表現と印象因子による印象推定の方法[14, 15]を用いる。イメージ語の座標値を本論文では印象値と呼ぶ。

イメージ推定処理では、はじめに、入力されたイメージ語と感性空間の因子を構成する形容詞対で共立共起表現を作成する。共立共起表現とは二つの形容詞または形容動詞をつなげて用いる表現のことである。因子を構成する形容詞または形容動詞を W_1 と W_2 、イメージ語を W_T とし、 W_1 と W_T の共立共起表現 F_1 、 W_2 と W_T の共立共起表現 F_2 を作成する。例として、イメージ語 W_T を「明るい」、 W_1 を「粗野な」、 W_2 を「優雅な」とした場合、それぞれの共立共起表現は

F_1 : 「粗野で明るい」OR「明るくて粗野な」

F_2 : 「優雅で明るい」OR「明るくて優雅な」

となる。

作成した共立共起表現 F_1 、 F_2 をキーワードとしてWeb検索を行い、共立共起表現の検索ヒット件数からイメージ語の印象値を算出する。まず、 F_1 の検索ヒット件数 H_1 、 F_2 の検索ヒット件数 H_2 をYahoo! 検索APIを利用して取得する。次に、 W_1 と W_T の印象の類似度を S_1 、 W_2 と W_T の印象の類似度を S_2 として、文献[14]によって定義されている式(1)と式(2)を用いて、類似度を算出する。

$$S_1 = 2.5^{\log_{10} H_1} \quad (1)$$

$$S_2 = 2.5^{\log_{10} H_2} \quad (2)$$

W_1 を -1.0 、 W_2 を $+1.0$ としたとき、 W_1 と W_2 を両極とする1次元軸上における W_T の座標値 a は文献[15]を参考に式(3)で定義し、算出する。

$$a = 2.0 \times \frac{S_1^{-1}}{S_1^{-1} + S_2^{-1}} - 1.0 \quad (3)$$

表4 気品因子の場合のヒット件数, 類似度, 座標値

形容詞	ヒット件数	類似度	座標値
粗野な	7	2.17	0.78
優雅な	1422	17.97	
なじみにくい	1	1	0.74
なじみやすい	117	6.65	
うっとうしい	57	5	0.46
すがすがしい	714	13.66	
下品な	43	4.47	0.81
上品な	13610	44.16	

以上の手順を因子を構成する形容詞対全てに対して行い, 各形容詞対に対するイメージ語の座標値を求める。

次に, 各形容詞対に対するイメージ語の座標値から, イメージ語の印象傾向を求める。上記の手順で求めたイメージ語の座標値が $[-1.0, 0.0]$ である形容詞対の数と, $[0.0, +1.0]$ である形容詞対の数を比較し, 数が多いほうの印象の傾向が強いと判断する。例としてイメージ語を「明るい」としたときの気品因子における印象傾向を求める場合, 表4に示すように, 4つの形容詞対に対する座標値が $[0.0, +1.0]$ の範囲にあることから, 「明るい」というイメージ語は「気品がある」傾向が強いと言える。

各形容詞対の中でイメージ語の印象傾向が強かった側の形容詞, 形容動詞とイメージ語の共立共起表現の検索ヒット件数が最も多かったもの, 2番目に多かったもののイメージ語の座標値を信頼できる値と考え, 座標値の平均をとる。平均をとった座標値がイメージ語の因子に対する印象値となる。イメージ語を「明るい」とした場合「気品がある」の傾向が強いことから, 検索ヒット件数が最も多い「上品な」に対する座標値0.81と, ヒット件数が2番目に多い「優雅な」に対する座標値0.78を平均し, 気品因子に対する印象値は0.79となる。活動性因子に対する印象値も同様に求められる。

2.4 ファジィ推論によるパラメータ決定

本論文で扱うイメージ語の印象値と演奏表情パラメータ値との関係は定性的であり, その表現にはあいまいさが含まれている。そこで, MUSAIでは, そのあいまいな関係をファジィルールで表現し, ファジィ推論を用いて演奏表情パラメータ値を求めることとする。イメージ推定処理で得られた印象値から, Mamdaniの方法[16]によるファジィ推論を行い, 演奏表情パラメータ値を求める。

メンバーシップ関数の定義

演奏表情パラメータ値の推論に用いるファジィルー

ルは, 前件部を活動性因子と気品因子に対するイメージ語の印象値, 後件部を演奏表情パラメータ値とした, 以下に示すような形とする。

if “活動性因子の印象値 a_A is LI”
and “気品因子の印象値 a_C is VIE”
then 基準テンポ $TempoBase$ is Allegro

図5, 図6はファジィルールの前件部の変数となる感性空間の各因子に対応するメンバーシップ関数である。

演奏表情パラメータに対応する後件部メンバーシップ関数を図7に示す。 $TempoBase$ に対応するメンバーシップ関数は図7(a)に示すように, 台形型メンバーシップ関数とし, メンバーシップ関数の値は一般的な機械式メトロノームに表記されているBPM (Beats Per Minute)¹と速度記号を参考に定義する。 $VelocityBase$ に対応するメンバーシップ関数はMIDI形式による楽曲データ作成の際に目安とされているベロシティ値を参考に図7(b)のように定義する。その他, $LengthSign$, $LengthBase$, $TempoVar$, $TempoRange$, $VelocityVar$, $VelocityRange$ はそれぞれ図7(c)から図7(f)のように定義する。表5に $TempoRange$, $VelocityRange$ に対応するメンバーシップ関数の値を示す。

図5 活動性因子に対応するメンバーシップ関数

図6 気品因子に対応するメンバーシップ関数

1 演奏のテンポを示す単位であり, 1分間に刻まれる拍の数でテンポを表す。

図7 演奏表情パラメータに対応する後件部メンバーシップ関数

表5 *TempoRange*, *VelocityRange* のメンバーシップ関数のパラメータ値

パラメータ	<i>TempoRange</i>	<i>VelocityRange</i>
<i>lw</i>	0	0
<i>sw</i>	0.15	0.175
<i>w</i>	0.3	0.35
<i>fw</i>	0.45	0.525
<i>vw</i>	0.6	0.7

	沈静的なイメージ	活性的なイメージ
テンポ	遅い だんだん遅く	速い だんだん速く
音の強弱	弱い だんだん弱く	強い だんだん強く
音の長短	長い	短い

図8 演奏表情とイメージの定性的な関係

ファジィルールの作成

本論文では、文献[17]による演奏表情の要素と演奏表情によって表されるイメージとの定性的な関係と、感性空間を構築するための実験で実験参加者に提示した演奏表情が持つ因子得点と演奏表情パラメータ値の重相関係数の二つを利用してファジィルールを作成する。

テンポ、音量の変化の仕方を示すパラメータ、演奏時に音を長くするか短くするかを示すパラメータである *TempoVar*, *VelocityVar*, *LengthSign* を推論するファジィルールは、ルール前件部に対応する感性空間の活動性因子と気品因子について、図8に示す演奏表情とイメージの定性的な関係[17]の活性的なイメージ、沈静的なイメージに因子の両軸を対応させて作成する。文献[17]によると、演奏の要素の状態によって図8に示す、活性的なイメージまたは沈静的なイメージを表すことができると言われている。図8の活性的なイメージには硬い、はやり、強いなどの印象が含ま

れ、沈静的なイメージは柔らかい、ゆっくり、弱いなどの印象が含まれると述べられていることから、本論文では、活動性因子と気品因子に含まれる形容詞の印象を考慮し、活動性因子の「活動的である」、気品因子の「気品がない」側を図8の活性的なイメージ、「活動的でない」「気品がある」側を沈静的なイメージに対応させ、*TempoVar*, *VelocityVar*, *LengthSign* それぞれについて、25個のファジィルールを作成する。このうち、*TempoVar* の値を推論するファジィルールを表6に示す。後件部メンバーシップ関数のラベルC, D, Iは、*TempoVar* が取りうる値である *constant*, *decrease*, *increase* に対応しており、それぞれ、テンポを一定に保つ、テンポをだんだん遅くする、テンポをだんだん速くすることを示す。

TempoBase, *VelocityBase*, *LengthBase*, *TempoRange*, *VelocityRange* の値を推論するファジィルールは、感性空間を構築する実験で実験参加者に提示した演奏表情の因子得点と演奏表情パラメータの重相関

表6 *TempoVar* のファジィルール

	VIE	LIE	N	LE	VE
VI	C	C	D	D	D
LI	C	C	D	D	D
N	I	I	C	D	D
LA	I	I	I	C	C
VA	I	I	I	C	C

表8 *TempoBase* のファジィルール

	VIE	LIE	N	LE	VE
VI	L	An	An	An	An
LI	An	An	An	An	M
N	M	M	M	M	Al
LA	Al	Al	Al	Al	Al
VA	Al	Al	Al	P	P

表7 相関の有無

	<i>TempoBase</i>	<i>VelocityBase</i>	<i>LengthBase</i>		<i>TempoVar</i>		<i>VelocityVar</i>	
			-1	1	increase	decrease	increase	decrease
相関の有無	あり	あり	あり	なし	あり	あり	あり	あり

係数を用いて作成する。2.3章で述べたイメージ推定処理で求められる印象値は、活動性因子、気品因子それぞれに対して、入力されたイメージ語がどのくらい似ているかを表す値である。また、感性空間を構築する実験で提示した演奏表情の因子得点は、活動性因子、気品因子に対してどのくらいの重みを持っているかを表す値である。そのため、本論文では、印象値と因子得点を同等のものとして扱うことを考える。このことから、 $[-1.0, 1.0]$ で正規化した演奏表情の活動性因子に対応する因子得点と気品因子に対応する因子得点の2つを重回帰分析の説明変数とし、目的変数を実験参加者に提示した演奏表情の *TempoBase*, *VelocityBase*, *LengthBase*, *TempoRange*, *VelocityRange* の値として重相関係数を求める。ここで *LengthBase* は *LengthSign* の値ごとに、*TempoRange* と *VelocityRange* はそれぞれ *TempoVar* と *VelocityVar* の値ごとに重相関係数を求める。表7に演奏表情の因子得点と演奏表情パラメータの相関の有無を示す。

表7より、*LengthSign* = 1 のときの *LengthBase* 以外のパラメータには相関が見られることから、重回帰式の説明変数に活動性因子、気品因子の値-1.0、-0.5、0.0、0.5、1.0をそれぞれ代入して求められる予測値の適合度が高いファジィ集合を後件部とすることで、ファジィルールを作成する。重回帰式の説明変数に代入する値は、図5と図6に示す活動性因子と気品因子に対応するファジィ集合のメンバーシップ値が1.0となる値である。本論文では *TempoBase*, *VelocityBase*, *LengthBase*, *TempoRange*, *VelocityRange* の値を推論するファジィルールがそれぞれ25個ずつ作成された。*TempoRange* と *VelocityRange* は、*TempoVar* と *VelocityVar* が *increase* であるか *decrease* であるかにより、ファジィルールが異なる。このうち、本論文で作成した *TempoBase* を推論するファジィルールを表8に示す。後件部メンバーシップ関数のラ

ベル L, An, M, Al, P はそれぞれ, *Largo*, *Andante*, *Moderato*, *Allegro*, *Presto* を省略して表記したものである。因子得点との相関が見られなかった *LengthSign* = 1 のときの *LengthBase* の値は、パラメータが取りうる値の中央値である 5.75 とし、定数として扱う。

2.5 ユーザとの対話による演奏表情の修正

提示された演奏表情にユーザが満足できなかった場合、演奏表情の修正を行う。ユーザは提示された演奏表情に満足できない場合に、演奏表情の印象をどのくらい変化させたいかを「少し元気な」「だいたい激しくない」というように、程度を表す副詞と、はじめに入力したイメージ語の肯定形もしくは否定形を組み合わせで入力する。本論文では、ユーザが演奏表情をどの程度修正したいかの表しやすさを考慮し、用いる副詞の数を3とする。イメージ語と組み合わせても違和感のない副詞として、文献[18]で挙げられている中から、本論文では「少し」「だいたい」「もっと」の3語を選択する。文献[18]では、これら3語の大小関係は、小さいほうから「少し」「だいたい」「もっと」となっているが、程度を表す副詞は国文学的な観点からの理解と、人間が日常生活で用いる際の理解とが必ずしも一致しない[19]という指摘があることから、本論文では、程度を表す副詞を用いた指示について、実験的に副詞の大小関係を明らかにしている文献[20]を参考に、小さいほうから「少し」「もっと」「だいたい」という大小関係でこの3語を用いることとする。

2.5.1 演奏表情の比較実験

本論文では、演奏表情の修正手法を検討するために、感性空間上で印象値が異なる2つの演奏表情の印象を相対的に評価する比較実験を行い、得られた結果の分析を行う。

図9 比較実験で用いる評価尺度の例

表9 実験で使用するイメージ語

柔かい	かたい	暖かい	冷たい
軽い	重い	輝かしい	かよわい
なめらかな	生気のない	健康的な	病的な
叙情的な	現実的な	幻想的な	悲観的な
楽観的な	大胆な	派手な	やぼったい
男性的な	女性的な	繊細な	

比較実験の内容

比較実験では、イメージ語を反映した演奏表情 A と、演奏表情 A とは異なる印象値から生成した演奏表情 B の 2 つの演奏表情を実験参加者に提示する。実験参加者は、演奏表情 A の印象に対して演奏表情 B の印象はどのくらい差があると感じるかを図9 に示す尺度を用いて評価する。例えば、イメージ語「明るい」が反映されている演奏表情 A と、異なる印象値から生成された演奏表情 B を比較し、B がより明るいと感じるか、明るくないと感じるか进行评估する。演奏表情 A に反映させるイメージ語は表9 に示す23 語である。イメージ語1 語につき、演奏表情 B を40 個用意する。演奏表情 B を生成するための印象値はランダムに決定される。比較実験の実験参加者は20 代の大学生、大学院生7 名である。

比較実験の結果

本論文では、図5、図6 に示したように感性空間の各因子に対応するファジィ集合が定義されている。そこで、演奏表情 A と演奏表情 B の印象値が各因子のどのファジィ集合に属するかを求め、ファジィ集合間の距離ごとに、各副詞に当てはまると評価された演奏表情の累積割合を図10 に示す。グラフ中、横軸の数値0 は演奏表情 A と演奏表情 B の印象値が同じ集合に属していることを示し、1 から4 は演奏表情 A の印象値から見て、演奏表情 B の印象値がいくつ隣の集合に属しているかを示す。図10 からは、副詞の程度が「少し」と、小さい場合にはファジィ集合間の距離が短いうちに累積割合が100% に近づいており、副詞の程度が「もっと」「だいたい」と大きくなるほど、ファジィ集合間の距離が大きくなるにつれて、累積割合が100% に近づいていることがわかる。これは、演奏表情 A との距離が近いほど、副詞の程度が小さいと評価され、距離が遠いほど、副詞の程度が大きいと評価されていると言える。また、距離ごとの累積割合につ

図10 距離ごとの各副詞に当てはまると評価された演奏表情の累積割合

いて、ライアンの方法による多重検定を行った結果、演奏表情 A と演奏表情 B が同じファジィ集合に属している場合、1 つ隣のファジィ集合に属している場合には、「少し」と評価された演奏表情 B と、「もっと」「だいたい」と評価された演奏表情 B の累積割合の間に有意水準5 %で有意差が見られた。そして、演奏表情 A と演奏表情 B が2 つ隣のファジィ集合に属している場合には、「少し」と評価された演奏表情 B と「もっと」と評価された演奏表情 B の累積割合の間に有意水準5 %で有意差が見られた。これらのことから、2 つの演奏表情の印象値の距離が大きくなると、演奏表情が与える印象の程度の差も大きくなると考えられる。

図11 にイメージ語「軽い」と「幻想的な」の場合の演奏表情 B の評価結果を感性空間上にプロットしたグラフを示す。グラフ中の●印は演奏表情 A の印象値であり、その他の記号は評価に応じた演奏表情 B の印象値である。グラフの上にした評価尺度は演奏表情 B に対する尺度であり、グラフ中の記号と対応する。また、×印は演奏表情 B がそのイメージ語に該当しないと評価されたことを示す。図11(a)では第1 象限、第4 象限の印象値から生成された演奏表情 B が、演奏表情 A よりも「軽い」という印象を感じると評価されているが、図11(b)では第1 象限、第4 象限の印象値から生成された演奏表情 B は、演奏表情 A よりも「幻想的な」とは反対の否定的な印象を感じると評価されていることがわかる。他のイメージ語についても「軽い」「幻想的な」と同様に、演奏表情 A の印象値から見た演奏表情 B の印象値の方向と印象の感じ方には規則性が見られないという結果が得られている。

比較実験の分析結果から、感性空間上での二つの演奏表情の印象値間の距離には程度副詞との関連が見られるが、二つの演奏表情の印象値間の方向性には印象

図 11 比較実験で提示した演奏表情の座標とその評価結果

の強弱との関連は見られないことがわかった。

2.5.2 演奏表情の修正

比較実験の分析結果から、修正指示語の程度副詞に応じて、感性空間上でユーザが望む演奏表情が修正前の演奏表情の印象値からどの程度離れているかは求められるが、どの方向にあるかを推測することは難しい。そのため、ユーザが望む演奏表情に修正するためには、修正した演奏表情が満足できるものであるかどうかをユーザに評価してもらう必要がある。

Interactive Particle Swarm Optimization

本論文では、生成された演奏表情がイメージ語どの程度反映しているかというユーザの評価とユーザが入力する修正指示語を感性空間上に反映しやすく、また、ユーザが満足する演奏表情の印象値に対応する感性空間上の座標値を求めやすいことから、群知能の一種であるParticle Swarm Optimization [21] (以下、PSO とする) の評価関数を人間の評価に置き換えた Interactive Particle Swarm Optimization [22] (以下、IPSO とする) を用いる。PSO, IPSO では、位置と速度を持つ個体群が解空間を移動し、最適な位置を探索する。各個体が持つ位置と速度は式(4)、式(5)で計算される。

$$\mathbf{x}_j(t+1) = \mathbf{x}_j(t) + \mathbf{v}_j(t) \quad (4)$$

$$\begin{aligned} \mathbf{v}_j(t+1) = & w(t)\mathbf{v}_j(t) + c_1r_1(\mathbf{p}_j(t) - \mathbf{x}_j(t)) \\ & + c_2r_2(\mathbf{g}(t) - \mathbf{x}_j(t)) \end{aligned} \quad (5)$$

$\mathbf{x}_j$ は個体 j の位置、 $\mathbf{v}_j$ はその速度を表す。式(5)の $\mathbf{p}_j$ は個体 j が t 世代目までで最もよい評価を得た位置、 $\mathbf{g}$ は個体群の中で最もよい評価を得た位置であ

る。 r_1 と r_2 はそれぞれ $[0.0, 1.0]$ の一様乱数、 w は慣性係数である。 c_1, c_2 は学習係数であり、 c_1 は各個体が最もよい評価を得た位置 $\mathbf{p}_j$ に近づく重み、 c_2 は個体群の中で最もよい評価を得た位置 $\mathbf{g}$ に近づく重みを表す。PSO, IPSO では一般的に係数 w, c_1, c_2 の値は任意に決定されることから、本提案手法では、ユーザが入力する修正指示語とユーザによる演奏表情の評価を w, c_1, c_2 と対応付ける。

提案手法でのパラメータ設定

本論文では、IPSO における個体が演奏表情に対応し、個体の位置 $\mathbf{x}_j$ が演奏表情 j の印象値を表す。任意に決められるパラメータ w, c_1, c_2 のうち、慣性係数 w を以下のように演奏表情ごとに設定する。

$$w_j(t) = |(\mathbf{h}_j(t) - \mathbf{l}_j(t)) \times \mathbf{adv}| \quad (6)$$

ここで、 $\mathbf{h}_j, \mathbf{l}_j$ はそれぞれ、 t 世代目の演奏表情 j の探索範囲の上限値と下限値であり、以下の式で表される。

$$\mathbf{h}_j(t) = \begin{pmatrix} h_{A_j}(t) \\ h_{C_j}(t) \end{pmatrix} \quad (7)$$

$$\mathbf{l}_j(t) = \begin{pmatrix} l_{A_j}(t) \\ l_{C_j}(t) \end{pmatrix} \quad (8)$$

$h_{A_j}(t), l_{A_j}(t)$ は活動性因子における演奏表情 j の探索範囲の上限値と下限値を示し、 $h_{C_j}(t), l_{C_j}(t)$ は気品因子における演奏表情 j の探索範囲の上限値と下限値を示す。また、 $\mathbf{adv}$ はユーザが入力する修正指示語の副詞に対応した値であり、2.5.1 節の比較実験の結果より、1 回修正した後の演奏表情は「少し」の場合にはひとつ隣のファジィ集合、「もっと」の場合には二つ隣の

図12 感性空間の活動性因子の軸における探索範囲の変化

ファジィ集合、「だいぶ」の場合には三つ隣のファジィ集合に属すると考える。2回目以降の修正では、探索範囲が狭まっていくと考えられるため、1回目の修正時のファジィ集合間の距離を参考にし、「少し」の場合は探索範囲の $\frac{1}{4}$ である0.25, 「もっと」の場合は $\frac{1}{2}$ である0.5, 「だいぶ」の場合は $\frac{3}{4}$ である0.75とする。例として、修正指示語「もっと明るい」から生成された t 世代目の演奏表情があるとき、 $(t+1)$ 世代目の演奏表情の印象値を求めるための、探索範囲の上限 $h_j(t+1)$ と下限 $l_j(t+1)$ を決定する手順を以下に示す。また、図12に活動性因子の軸を例として、以下の手順を施した場合の感性空間の探索範囲の推移を示す。図12の l_{A_j} , h_{A_j} は活動性因子軸についての探索範囲の下限値と上限値を表し, x_{A_j} は活動性因子軸についての印象値を表す。なお、探索範囲の上限 $h_j(t+1)$ と下限 $l_j(t+1)$ は感性空間の気品因子の軸についても同様の手順で決定される。

- t 世代目の j 番目の演奏表情に対するユーザーの評価が肯定的だった場合

－ t 世代目の修正指示語の形容詞が $(t-1)$ 世代目と同じ: 「少し・もっと・だいぶ」+「明るい」

- * t 世代目の演奏表情の印象値が、 $(t-1)$ 世代

目の演奏表情の印象値より大きい(図12(a))

$$\cdot l_j(t+1) = x_j(t-1)$$

$$\cdot h_j(t+1) = x_j(t)$$

- * t 世代目の演奏表情の印象値が、 $(t-1)$ 世代目の演奏表情の印象値より小さい(図12(b))

$$\cdot l_j(t+1) = l_j(t)$$

$$\cdot h_j(t+1) = x_j(t)$$

－ t 世代目の修正指示語の形容詞が $(t-1)$ 世代目と異なる: 「少し・もっと・だいぶ」+「明るくない」

- * t 世代目の演奏表情の印象値が、 $(t-1)$ 世代目の演奏表情の印象値より大きい(図12(c))

$$\cdot l_j(t+1) = x_j(t-1)$$

$$\cdot h_j(t+1) = x_j(t)$$

- * t 世代目の演奏表情の印象値が、 $(t-1)$ 世代目の演奏表情の印象値より小さい(図12(d))

$$\cdot l_j(t+1) = x_j(t)$$

$$\cdot h_j(t+1) = x_j(t-1)$$

- t 世代目の j 番目の演奏表情に対するユーザーの評価が否定的だった場合

– t 世代目の演奏表情の印象値が、 $(t-1)$ 世代目の演奏表情の印象値より大きい(図12(e))

$$\begin{aligned} * l_j(t+1) &= l_j(t) \\ * h_j(t+1) &= x_j(t) \end{aligned}$$

– t 世代目の演奏表情の印象値が、 $(t-1)$ 世代目の演奏表情の印象値より小さい(図12(f))

$$\begin{aligned} * l_j(t+1) &= x_j(t) \\ * h_j(t+1) &= h_j(t) \end{aligned}$$

肯定的な評価を得た t 世代目の演奏表情に対する修正指示語の形容詞が $(t-1)$ 世代目の形容詞と同じであった場合、 $(t-1)$ 世代目の印象値から t 世代目の印象値へ方向と同じ向きに $(t+1)$ 世代目の印象値があると、図12(a)、図12(b) に示される探索範囲が求められる。また、 t 世代目の演奏表情に対する修正指示語の形容詞が $(t-1)$ 世代目とは異なった場合、 $(t-1)$ 世代目の印象値から t 世代目の印象値へ方向とは逆の向きに $(t+1)$ 世代目の印象値があると、図12(c)、図12(d) に示される探索範囲が求められる。 t 世代目の演奏表情が否定的な評価を得た場合、修正指示語の形容詞によらず、 $(t-1)$ 世代目の印象値から t 世代目の印象値へ方向と同じ向きには $(t+1)$ 世代目の印象値はないとし、図12(e)、図12(f) に示される探索範囲となる。

式(5)の学習係数 c_1 、 c_2 は以下のように設定する。

$$c_2 = \frac{1}{1 + e^{-(EvalP(t)-5)}} \quad (9)$$

$$c_1 = 1.0 - c_2 \quad (10)$$

PSO、IPSO では c_1 と c_2 の値は1.0 前後が適切であることが経験則から言われていることから、式(9)、式(10)から求められた c_1 、 c_2 を $[0.5, 1.5]$ の範囲に変換する。式(9)の変数 $EvalP(t)$ は、 t 世代目において、ユーザからの評価が肯定的だった演奏表情の数を表す。評価が肯定的だった演奏表情数 $EvalP(t)$ の値が5 以上の場合には、式(9)から c_2 は1 に、 c_1 は0 に近づき、式(5) から $(t+1)$ 世代目の個体は t 世代目で演奏表情全体で最も良い評価を得た演奏表情の位置 g へ移動しやすくなり、探索範囲が狭まる。一方、 $EvalP(t)$ が5 以下の場合には、式(9)から c_2 は0 に、 c_1 は1 に近づき、式(5)から $(t+1)$ 世代目の個体は、これまでに各演奏表情が最も良い評価を得た位置 p_j へ移動しやすくなり、ユーザの印象を反映すると思われる演奏表情をさらに探索する。

演奏表情の修正手順

生成した演奏表情にユーザが満足できない場合、程

度副詞「少し」「もっと」「だいぶ」と肯定形もしくは否定形のイメージ語を組み合わせた修正指示語を入力する。本論文で提案する修正手法では、演奏表情の1 回目の修正と2 回目以降の修正で手順が異なる。

(1) 1 回目の修正

定義したファジィ集合に従って感性空間を図11に示すような25 個の領域に分割する。比較実験の結果から、程度副詞が大きくなると修正前の演奏表情の印象値から遠い領域に修正後の演奏表情の印象値があると考えられるため、入力された修正指示語の副詞が「少し」の場合には修正前の演奏表情の印象値の隣の領域、「もっと」の場合には2 つ隣の領域、「だいぶ」の場合には3 つ隣の領域が修正後の演奏表情の印象値がある領域であると仮定する。そして、修正指示語の副詞に対応する各領域内でランダムに印象値を決定する。これを修正した演奏表情の印象値とし、位置 x とする。また、演奏表情の速度 v は0 とする。求められた印象値から演奏表情が生成され、修正した演奏表情としてユーザへ提示される。1 回目の修正でユーザに提示される演奏表情の数は、修正前の演奏表情の印象値と修正指示語の程度副詞に依存し、3 個から16 個の演奏表情がユーザへ提示されることとなる。提示される演奏表情の数が10 個に満たなかった場合には、ランダムな印象値を持つ演奏表情を生成し、ユーザに提示される演奏表情の数が10 個を満たすようにする。演奏表情の数が10 個を超えていた場合には、生成された全ての演奏表情がユーザに提示される。修正された演奏表情に対して、ユーザは印象を反映しているかどうかを以下の尺度で評価する。

- 1：反映していない
- 0：どちらでもない
- +1：やや反映している
- +2：反映している

提案手法で用いる IPSO は、 t 世代目までで良い評価を得た個体の方向に向かって $(t+1)$ 世代目の個体が移動するアルゴリズムである。式(5) の p_j は、演奏表情 j が t 世代目までで最もよい評価を得た位置であることから、 $(t+1)$ 世代目の演奏表情の位置を p_j の方向に移動させるために、印象を反映している場合の評価を2 段階とし、印象を反映していない場合の評価を1 段階と設定する。そして、修正した演奏表情の中で最も印象を反映しているものをベスト演奏表情として選択する。修正された演奏表情について満足できない場合には、ユーザは再び修正指示語を入力する。

(2) 2 回目以降の修正

2 回目以降の修正からはユーザに提示される演奏表情の数を10 個とする。1 回目の修正において、10 個

以上の演奏表情が生成されていた場合には、ユーザの評価が高い10個を以降の修正で引き継ぐ。

前世代の演奏表情に対するユーザの評価と、入力された修正指示語から、式(4)から式(10)を用いて演奏表情の位置 x_j と速度 v_j を更新し、演奏表情を修正する。

以上の手順をユーザが満足できる演奏表情が生成されるまで繰り返す。

3. 実験

提案したMUSAIの有効性を検証するための実験を行う。本論文では、MUSAIから修正手法を除いた生成のみ行うシステムと、MUSAIの2つを用いて、実験参加者が演奏表情を生成し、評価する。

3.1 修正を除いたシステムの評価実験

実験参加者は表9に示すイメージ語をひとつ選択し、システムを用いて表10に示す4曲について、それぞれ演奏表情を生成する。そして、生成された演奏表情について、表11に示す評価項目を-3から+3の7段階の尺度で評価する。なお、イメージ語の選択は、楽曲が持つ印象に影響を受けないようにするため、参加者は楽曲を聞かない状態で行う。実験参加者は19歳から25歳の男女20名である。

実験結果と考察

表12にこの実験が入力されたイメージ語を示す。図13に実験参加者の評価の平均値と、母平均から推

図13 評価項目に対する実験参加者の評価の平均値と95%信頼区間

定した95%信頼区間のグラフを示す。評価項目Q1を含めた全ての項目について、評価の平均値が全て0.5以上であり、肯定的な評価が得られている。したがって、システムにより、演奏表情に実験参加者の印象をほぼ反映できていることがわかる。一方で、95%信頼区間の下限値が0.5に近い評価項目や0.5を下回っている評価項目があり、これらは実験参加者20名のうち4名が、システムで生成された演奏表情に対して否定的な評価をしていたため、95%信頼区間の下限値がこのような値になったと考えられる。システムで用いている感性空間は多くの実験参加者のデータから構築されており、ファジィルールは重回帰分析により求められた演奏表情の要素とイメージの関係に基づいて求められている。そのため、感性空間とファジィルールは形容詞、形容動詞で表される印象に対するユーザ自身の主観を必ずしも反映できるとは限らないと言える。従って、生成された演奏表情について全ての実験参加者が肯定的な評価をするとは限らないと考えられる。

3.2 MUSAIの評価実験

実験参加者は表10に示す4曲について、演奏表情に反映させたいイメージ語を自由に入力する。そして、満足する演奏表情が生成されるまで、演奏表情の修正を繰り返す。ただし、修正の上限回数は10回とし、10回目の修正による演奏表情が生成された時点で実験終了とする。完成した演奏表情に対する評価項目を表13に示す。評価項目のうち、Q5からQ8では評価の理由についても自由記述で回答してもらう。Q5からQ8の評価尺度では、否定的な評価と肯定的な評価で用いている副詞が異なっているが、実験時に実験参加者に対して、日本語として違和感を与えないことを考慮したためである。また、同様の評価尺度はアン

表10 実験で使用する楽曲

楽曲名	作曲者名
グノシエンヌ 第1番	エリック・サティ
エコセーズ ト長調 WoO.23	L.v. ベートーヴェン
華麗なる大円舞曲 変ホ長調 Op.18	F.F. ショパン
ピアノソナタ第2番 変ロ短調 Op.35-3	F.F. ショパン

表11 演奏表情の評価項目

	評価項目
Q1	入力したイメージ語を反映した演奏表情を生成することができましたか？
Q2	演奏全体のテンポは適切でしたか？
Q3	テンポの変化の仕方は適切でしたか？
Q4	テンポの変化の程度は適切でしたか？
Q5	演奏全体の音量は適切でしたか？
Q6	音量の変化の仕方は適切でしたか？
Q7	音量の変化の程度は適切でしたか？
Q8	音の長さの変化（短い・長い）は適切でしたか？
Q9	音の短さ・長さは適切でしたか？

表12 実験参加者が入力したイメージ語

実験参加者	イメージ語			
	楽曲番号 1	楽曲番号 2	楽曲番号 3	楽曲番号 4
A	派手な	大胆な	幻想的な	暖かい
B	かわいい	悲観的な	病的な	大胆な
C	軽い	派手な	かたい	暖かい
D	派手な	病的な	輝かしい	楽観的な
E	冷たい	病的な	幻想的な	輝かしい
F	悲観的な	輝かしい	病的な	幻想的な
G	幻想的な	派手な	大胆な	柔らかな
H	病的な	男性的な	女性的な	楽観的な
I	派手な	健康的な	大胆な	病的な
J	幻想的な	女性的な	暖かい	生気のない
K	女性的な	幻想的な	—	大胆な
L	暖かい	なめらかな	かたい	派手な
M	やぼったい	男性的な	暖かい	かたい
N	派手な	かたい	重い	輝かしい
O	なめらかな	大胆な	幻想的な	冷たい
P	派手な	重い	輝かしい	輝かしい
Q	楽観的な	楽観的な	輝かしい	輝かしい
R	輝かしい	重い	健康的な	病的な
S	かたい	軽い	派手な	重い
T	幻想的な	男性的な	かたい	生気のない

表13 評価項目と評価尺度

No.	評価項目と評価尺度		
Q1	入力したイメージ語を反映した演奏表情を生成することができましたか？		
	(1) とても不満足	(5) やや満足	
	(2) 不満足	(4) どちらでもない	(6) 満足
	(3) やや不満足	(7) とても満足	
Q2	演奏表情のテンポは、イメージ語に対して適切でしたか？		
Q3	演奏演奏の音量は、イメージ語に対して適切でしたか？		
Q4	演奏演奏の音の長さは、イメージ語に対して適切でしたか？		
	(1) とても不適切	(5) やや適切	
	(2) 不適切	(4) どちらでもない	(6) 適切
	(3) やや不適切	(7) とても適切	
Q5	演奏表情が、入力したイメージ語に近づいていくように感じましたか？		
Q6	テンポが、入力したイメージ語に近づいていくように感じましたか？		
Q7	音量が、入力したイメージ語に近づいていくように感じましたか？		
Q8	音の長さが、入力したイメージ語に近づいていくように感じましたか？		
	(1) まったく感じなかった	(5) やや感じた	
	(2) 感じなかった	(4) どちらでもない	(6) 感じた
	(3) あまり感じなかった	(7) とても感じた	

ケート調査でよく用いられている[23]ことから、この評価尺度を採用している。実験には20代の大学生、大学院生12名が実験参加者として参加する。

実験結果と考察

生成実験の結果、48個の演奏表情が生成され、そのうち45個が修正を経て生成された演奏表情であった。表14にこの実験で入力されたイメージ語を示す。また、図14に修正を経て完成した演奏表情についての評価項目Q1からQ4に対する評価値の分布を示す。グラフの縦軸の番号は表13の評価項目の番号に対応しており、グラフ中の値は表13の評価値に対応している。完成した演奏表情に入力したイメージ語が反映されて

表14 実験で入力されたイメージ語

実験参加者	イメージ語			
	楽曲番号 1	楽曲番号 2	楽曲番号 3	楽曲番号 4
A	豪快な	楽しげな	熱い	理不尽な
B	重い	かわいい	なめらかな	悲観的な
C	明るい	寂しい	眠い	機械的な
D	楽しい	柔らかな	物悲しい	激しい
E	生気のない	かたい	軽い	優しい
F	悲観的な	軽い	楽観的な	病的な
G	軽やかな	弱い	激しい	おとなしい
H	派手な	明るい	男性的な	軽い
I	軽い	大胆な	病的な	女性的な
J	緩やかな	優しい	元気な	豪快な
K	軽快な	不思議な	かつこいい	可愛い
M	明るい	楽観的な	重い	女性的な

図14 評価項目 Q1 から Q4 までの評価値の分布

いたかどうかの評価の割合は、評価項目Q1からQ4のいずれの評価項目においても、5以上の評価値が80%以上である。修正を経て生成された45個のうち39個が印象を反映していたことから、MUSAIにより、演奏表情とその要素であるテンポ、音量、音の長さには各実験参加者の印象が十分に反映されていることがわかる。

図15は、完成した演奏表情が満足だったかどうかを評価する評価項目Q1の結果と、完成した演奏表情が何度の修正を経て生成されたものであるかを示すグラフである。修正を経て完成した45個の演奏表情のうち、40個は1回から5回の修正によって生成されていることがわかる。実験では演奏表情の修正の上限回数を10回に制限したが、完成した演奏表情の修正回数から、5回までの修正で実験参加者の印象を十分に反映できていると言える。

肯定評価であった演奏表情について、完成するまでに入力された修正指示語の副詞の頻度を図16に示す。肯定評価であった演奏表情39個はそれぞれ、1回から7回、9回の修正が施されている。修正回数が6回より多い演奏表情はそれぞれ1個ずつであるため、5

図15 完成した演奏表情の修正回数と評価項目Q1に対する評価

図16 評価項目Q1で肯定評価を得た演奏表情の修正時に入力された副詞の頻度

回までの修正で入力された副詞の頻度に注目する。5回までの修正で入力された副詞の頻度を見ると、「少し」は修正回数が増えるにつれて入力される頻度が少なくなる傾向にあり、「だいぶ」は多くなる傾向にあることがわかる。「もっと」については、ほぼ横ばいであると言える。このことから、修正回数が1回から2回であった演奏表情は、最初に生成された演奏表情が印象をほぼ反映しており、修正の程度は小さかったと考えられる。また、修正回数が3回から5回であった演奏表情は、最初に生成された演奏表情からの修正の程度が大きかったと考えられる。否定評価であった演奏表情が完成するまでに入力された修正指示語の副詞の頻度を図17に示す。否定評価であった演奏表情4個はそれぞれ、1回、5回、10回の修正が施されている。そのうち、1回、5回の修正による演奏表情は、どちらも程度副詞「だいぶ」のみが修正時に入力されている。10回の修正による演奏表情は2個あり、1個は1回、5回と同様に「だいぶ」のみが修正時に入力されており、もう1曲は「もっと」が10回の修正のうち7回、「だいぶ」が3回入力されている。このこと

図17 評価項目Q1で否定評価を得た演奏表情の修正時に入力された副詞の頻度

から、否定評価だった演奏表情4個は、実験参加者が思い描いていた印象からかけ離れており、修正指示語、修正時の評価をうまく取り入れることができなかったと考えられる。

肯定評価だった演奏表情のうちの2個について、図18に感性空間上での演奏表情の座標の遷移を示す。グラフ中の◆は、最初に生成された演奏表情の座標を示し、値は修正して提示された各演奏表情に対する実験参加者の評価を示す。図18(a)に示す実験参加者Aが生成した演奏表情の座標の遷移を見ると、1回目の修正では活動性因子が0以上の演奏表情に対して肯定的な評価がされており、2回目の修正では、1回目の修正でよい評価が得られた演奏表情の方角に演奏表情が移動している。その後、3回目の修正では、2回目の修正で最もよい評価が得られた演奏表情(図中でBestと記載されている演奏表情)の方向に演奏表情が移動していることがわかる。また、図18(b)に示す実験参加者Bが生成した演奏表情の座標の遷移でも、実験参加者Aの場合と同様に、2回目の修正では演奏表情が、1回目の修正でよい評価を得られた演奏表情の座標の周辺に移動しており、3回目の修正においても、2回目の修正で最もよい評価が得られた演奏表情の周辺に座標が集まっていることがわかる。このことから、本論文で提案した修正手法によって、ユーザが入力した印象を反映する演奏表情が生成されていると言える。また、実験参加者Aと実験参加者Bの演奏表情の座標の遷移を比較すると、実験参加者A、実験参加者Bはどちらも同じイメージ語を演奏表情に反映させており、イメージ推定処理で求められた「明るい」の印象値は第1象限であったが、どちらも「明るい」印象を反映した演奏表情の座標は第4象限の範囲にあることが修正過程の評価からわかる。だが、実験参加者Aが感じる「明るい」を反映した演奏表情は活動性因子の値が1.0に近いのに対し、実験参加者Bの場合は気品因子の値が1.0に近いほど評価がよく、一方で活動性因子の値が1.0に近い演奏表情への評価は悪いことから、同じイメージ語を反映させた演奏表情

(a) 実験参加者 A が生成した演奏表情の座標の遷移

(b) 実験参加者 B が生成した演奏表情の座標の遷移

図 18 「明るい」印象を反映した演奏表情の座標の遷移

でも実験参加者によって印象値が異なることがわかる。
次に、MUSAI を用いて演奏表情を修正することで、演奏表情が入力したイメージ語に近づいたと感じ

たかどうかについての評価項目 Q 5 から Q 8 に対する評価値の分布を図 19 に示す。グラフの縦軸の番号は表 13 の評価項目の番号に対応しており、グラフ中の

図19 評価項目 Q5 から Q8 までの評価値の分布

値は表13の評価値に対応している。評価項目 Q5 から Q8 に対する評価結果から、修正を経て完成した演奏表情のうち、60%から80%が肯定的な評価を得ていることがわかる。このことから、MUSAI により演奏表情を生成する過程で、演奏表情が入力したイメージ語に近づいていくことを実験参加者が知覚できていると言える。一方で、少数ではあるが、生成実験において否定的な評価を得た演奏表情は、修正を重ねてもイメージ語が反映されなかった、同じような演奏表情が提示された、という実験参加者からの回答が見られた。

4. おわりに

本論文では、ユーザを指導者、システムを演奏者として見立て、実際の演奏における指導者と演奏者のやり取りのように、言葉で表された印象が入力されると、その印象を反映した演奏表情を生成し、さらに「もっと明るい」「少し元気でない」といった、程度を表す副詞と形容詞・形容動詞を用いて表される修正の指示から演奏表情を修正するシステムMUSAIについて述べ、評価実験の結果から、MUSAI と対話して演奏表情を修正していくことで、実験参加者それぞれが自由に入力した印象を反映した演奏表情を生成でき、提案した修正手法が有効に働いていることがわかった。また、提案した修正手法により、5回までの修正で、入力された印象を十分に反映した演奏表情を生成できていることがわかった。さらに、演奏表情の修正の過程において、実験参加者自身が修正によって演奏表情に印象が反映されていくことを知覚していることもわかった。生成実験において実験参加者が生成した各世代の演奏表情の座標の遷移、演奏表情に対する評価の遷移を分析したところ、参加者の演奏表情に対する評価と修正指示語が座標値の遷移に適切に反映されていること、入力された印象が同じであっても、参加者によ

って感性空間上で表される座標値は異なっていることがわかった。一方で、少数ではあるが、修正を重ねていくことで生成される演奏表情がイメージ語を反映しなくなる、似たような演奏表情が生成されていた。また、各世代で入力される修正指示語や提示された演奏表情に対する評価の反映には2世代から3世代の時間がかかっていたこともわかった。そのため、演奏表情の修正においてイメージ語を反映しなくなる、似たような演奏表情ばかりが生成されることを避けるために、例えば遺伝的アルゴリズムにおける突然変異のような仕組みを取り入れる、また、ユーザによって入力される修正指示語や演奏表情への評価をより少ない世代で反映させることで、MUSAI との対話の回数をさらに減らすことができるといった改善が考えられる。

参考文献

- [1] 保科洋. 生きた音楽表現へのアプローチ—エネルギーに基づく演奏解釈法. 音楽之友社, 1998.
- [2] 松原正樹, 深山覚, 奥村健太, 寺村佳子, 大村英史, 橋田光代, 北原鉄朗. 創作過程の分類に基づく自動音楽生成研究のサーベイ. コンピュータソフトウェア, Vol.30, No.1, 2013.
- [3] Alexis Kirke and Eduardo Reck Miranda. A survey of computer systems for expressive music performance. *ACM Computing Surveys*, Vol.42, No.1, pp.1-41, 2009.
- [4] Roberto Bresin and Anders Friberg. Emotional Coloring of Computer-Controlled Music Performances. *Computer Music Journal*, Vol.24, No.4, pp.44-63, 2000.
- [5] Steven R. Livingstone, Ralf Mühlberger, Andrew R. Brown, and Andrew Loch. Controlling musical emotionality: an affective computational architecture for influencing musical emotions. *Digital Creativity*, Vol.18, No.1, pp.43-53, 2007.
- [6] Sergio Canazza, Giovanni De Poli, Antonio Rodà, and Massimiliano Barichello. CaRo 2.0: A system for social active listening and expressive performing of music content. In *proc. of Rencon*, 2011.
- [7] 馬場隆, Sergio Canazza, Antonio Rodà, 片寄晴弘, Giovanni De Poli. テンポ・音量・音色の実時間制御を目的とした表情付けシステム. 情報処理学会研究報告. EC, エンタテインメントコンピューティング, 第2013巻, pp.1-5. 一般社団法人情報処理学会, 2013.
- [8] Sergio Canazza, Giovanni De Poli, Antonio Rodà, Alvis Vidolin. An abstract control space for communication of sensory expressive intentions in music performance. *Journal of New Music Research*, Vol.32, No.3, pp.281-294, 2003.
- [9] 秋口俊輔. ソフトコンピューティング手法を用いた曲印象からの楽曲自動生成システムの構築. 知能と情報: 日本知能情報ファジィ学会誌, Vol.21, No.5, pp.782-791, 2009.
- [10] MusicXML for Exchanging Digital Sheet Music. <http://www.musicxml.com/>.

- [11] 山脇一宏, 椎塚久雄. コレスポンデンス分析による音楽の特徴認識. 日本感性工学会 第20回あいまいと感性研究部会, pp.94-97, 2008.
- [12] Charles Egerton Osgood, George J. Suci, and Percy H. Tannenbaum. *The measurement of meaning*. University of Illinois Press, 1957.
- [13] Robert M. Groves, Floyd J. Fowler, Mick Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 調査法ハンドブック. 朝倉書店, 2011. 氏家豊, 大隅昇, 松本渉, 村田磨理子, 嶋真紀子 訳.
- [14] 清水浩平, 萩原将文. 形容詞共起を用いた単語の印象推定法. 電子情報通信学会論文誌 D, Vol.89, No.11, pp.2483-2490, 2006.
- [15] 盛多亮, 鬼沢武久. 形容詞・形容動詞の印象を用いたフレーズアニメーションの生成. 日本知能情報ファジィ学会誌, Vol.22, No.1, pp.121-134, 2010.
- [16] 本多中二, 大里有生. ファジィ工学入門. 海文堂, 1989.
- [17] Hans-Peter Schmitz. 演奏の原理. シンフォニア, 1977.
- [18] 田和真紀子. 程度副詞の評価性をめぐって. 宇都宮大学教育学部紀要. 第1部, Vol.61, pp.25-36, 2011.
- [19] 織田揮準. 日本語の程度量表現用語に関する研究. 教育心理学研究, Vol.18, No.3, pp.166-176, 1970.
- [20] Satoru SHIBATA, Tomonori YAMAMOTO, and Mitsuru JINDAI. Velocity Adjustment of Robots using Degree Adverb Based on Human Image. *The Japanese Journal of Ergonomics*, Vol.48, No.6, pp.324-333, 2012.
- [21] R. C. Eberhart and J. Kennedy. A new optimizer using particle swarm theory. In *Proc. of the 6th International Symposium on Micromachine and Human Science*, pp.39-43, 1995.
- [22] J. Madar, J. Abonyi and F. Szeifert. Interactive Particle Swarm Optimization. pp.314-319, 2005.
- [23] 豊田秀樹. 調査法講義. 朝倉書店, 1998.
(2014年10月24日 受付)
(2015年5月21日 採録)

【問い合わせ先】

〒084-0916 北海道釧路市大楽毛西2丁目32番1号
 釧路工業高等専門学校 情報工学科
 鈴木 未央
 TEL: 0154-57-7359
 E-mail: mio@kushiro-ct.ac.jp

著者紹介

すずき みお
鈴木 未央 [正会員]

2008年3月富山大学工学部卒業.
 2010年3月筑波大学大学院システム
 情報工学研究科博士前期課程修了.
 2014年3月同大学院同研究科博士
 後期課程修了. 現在, 釧路工業高等
 専門学校情報工学科助教. 博士(工学).
 人間とコンピュータとの対話による人
 間の感性理解と教育支援への応用の研
 究に従事. IEEE, 日本知能情報ファ
 ジィ学会, 各会員.

おにさわ たけひさ
鬼沢 武久 [正会員]

1975年3月東京工業大学制御工学
 科卒業. 1977年3月同大学院理工
 学研究科制御工学専攻修了. 現在, 筑
 波大学システム情報系教授. 工学博士
 (東京工業大学). 人間の主観, 判断,
 認識, 感性, フィーリングを処理する
 主観・感性コンピューティングに関心
 があり, インタラクティブ・デザイン
 に関する研究に従事. 1987年日本人
 間工学会から橋本賞, ISIS2005でOut-
 standing Paper Award, 2012年日本感
 性工学会から技術研究賞など. SOFT
 関係では第9期会長, 第13期評議員
 長などを務める.

Musical Expression Generation Reflecting User's Impression by Interaction

by

Mio SUZUKI and Takehisa ONISAWA

Abstract :

This paper describes an MUSAI (MUSical expression generation system by Adjective with Interaction) reflecting user's impression represented by an adjective, i.e., an image word. The MUSAI uses a Kansei space and fuzzy rules that represent a relationship between musical expression and impression. The inputted user's impression is mapped into a Kansei space and the parameter values of musical expression are obtained by fuzzy inference. Then, the musical expression is generated based on the obtained parameter values and the inputted MusicXML. If a user is not satisfied with generated musical expression, a user inputs a modification word with an adverb and an image word such as *more bright*, and the musical expression is modified by changing the coordinates values in a Kansei space. The validity of the MUSAI is confirmed by experiments. From the experimental results, it is found that the MUSAI successfully generates musical expressions reflecting each experiment participant's impression.

Keywords : musical expression, impression, Kansei space, fuzzy inference, interaction

Contact Address : Mio SUZUKI

Department of Information Engineering, National Institute of Technology, Kushiro College

2-32-1, Otanoshike-nishi, Kushiro, Hokkaido, 084-0916, Japan

TEL : 0154-57-7359

E-mail : mio@kushiro-ct.ac.jp