

ピアノ弾き語りアレンジのための音楽的緊張モデルの最適化 手法の提案

樋口 梨花^{1,a)} 竹川 佳成^{1,b)} 平田 圭二^{1,c)}

概要：本研究は、ピアノ弾き語り譜の生成に応用可能な、音楽的緊張の予測モデルの構築を目指す。我々は、音楽を聴いているときに無意識に様々な予測をしている。次の展開を予測できたとき、人間は緊張状態になり、予測が正しいのか否かを考える。予測通りの展開であった場合は緊張感から解放され、予測が裏切られた場合は緊張感が増幅する。このときの緊張感がのちに新たな情動を生むとされ、緊張感が大きいほど解消されたときの情動は大きくなる。音楽によって情動が生まれる仕組みには、この予測と裏切りによって引き起こされる音楽的緊張が関係している。一方、ピアノアレンジについての研究は盛んであるが、編曲後の楽曲の雰囲気維持されていないなど、アレンジの質について十分に考慮できていないという問題がある。これに対し本研究は、ピアノ弾き語りアレンジの楽譜の生成を行う際に、音楽的緊張を考慮することでアレンジの質の向上を目指す。本論文では、楽曲から得られる特徴量と、被験者実験によるデータから、その楽曲の音楽的緊張を予測するモデルを構築する手法を提案する。

Proposal of Optimization Method of Musical Tension Model for Piano Accompaniment for Singing Song

RIKA HIGUCHI^{1,a)} YOSHINARI TAKEGAWA^{1,b)} KEIJI HIRATA^{1,c)}

1. はじめに

音楽によって、人は強い緊張感や安静感、解消感を得ることがある。人は音楽の構造を知覚し、情動的な反応を起こしているとされるが、その仕組みは未だ解明されていない。現在、様々なアプローチによって、音楽と人の認知の関係を解明しようとする研究は多く存在する。その関係を解くための1つのアプローチとして「予測と裏切り」がある [1]。我々は、音楽を聴いているときに無意識に様々な予測をしている。次の展開を予測できたとき、人間は緊張状態になり、予測が正しいか間違いかを無意識に考える。その緊張感が、のちに新たな情動を生む。音楽によって引き起こされる情動は、予測が多少なりとも裏切られたことによって生まれる。例えば、曲やフレーズが終わるときの

コード進行は、和音を不安定なものから比較的安定したものへと進行させることで、曲が終わったという印象をつける役割がある。多くの人は過去の経験から、曲を評価するための「枠組み」を形作っている。その枠組みによって、無意識のうちに、曲の終わりの和音は安定性の高い和音（トニック）になると予測をしている。もし、トニックがすぐに聞こえてきた場合は、「次の展開は予測通りの展開なのか、期待を裏切る展開なのか」という緊張感は生まれず、感情を動かされることもない。しかし、予測を裏切るようなコード進行が続き、緊張感が増幅された上でトニックが聞こえてきた場合は、得られる情動がより大きくなる。コード進行以外にも、曲の終盤にテンポを意図的に遅くするというテクニックによって緊張感を生じさせることができる。このように、次の展開を仄めかしその通りに展開する、予測を裏切るという仕組みによって、緊張と弛緩は音楽の中で交互に繰り返される。

音楽的緊張は、音楽構造と人の認知の間に存在する概念である。音楽構造によって作られた緊張・弛緩状態は人の

¹ 公立はこだて未来大学
Futute University Hakodate

a) g2118031@fun.ac.jp

b) yoshi@fun.ac.jp

c) hirata@fun.ac.jp

認知に影響を及ぼす。音楽的緊張は、その影響度を測るための尺度となる。音楽的緊張はその曲のコード進行、テンポ、リズムといった特徴量によって作り出されている。それらの特徴量を意図的に操作することによってより緊張感の強い曲やフレーズを作り出すことが可能である。音楽的緊張について研究することは、音楽と人の認知の関係を探るうえで重要なアプローチである。

音楽的緊張の応用として、本研究は音楽の展開を考慮したピアノ弾き語り譜の生成を考えている。ピアノ弾き語りを行うためには、好みの楽曲がピアノ弾き語り用にアレンジされた譜面を入手する必要があるが、欲しい譜面が必ずしも入手できるとは限らない [13] [14]。一方で、アレンジされた譜面がない場合は、原曲か、構造のよく似ているピアノソロ用の譜面をもとに編曲を行うことがある。しかし、編曲には多大な労力と専門知識を必要とするため、楽譜の生成や編曲支援についての研究は常に需要がある。また、ピアノアレンジについての研究は盛んに行われているが、編曲後の楽曲の雰囲気が維持されていないなど、アレンジの質について、十分に考慮できていないという問題がある。例えば、楽曲中に同じフレーズが何度か出現したとしても、出現するタイミングや構成音の数などによって人の認知する緊張感は異なると考えられる。しかし、従来の手法ではこのような違いを考慮しきれていない。本研究では、音楽的緊張を用いることで、アレンジの質を高めることが可能であると考えられる。

本論文では、楽曲の各特徴量から音楽的緊張の予測モデルを作成する手法を提案する。従来の研究は、クラシック音楽を対象として、音楽的緊張に影響度の強い特徴量を使用していた。本研究で扱う楽曲は一般的な音楽を対象としている。一般的な音楽はクラシック音楽と異なる構成であるため、特徴量を見直す必要がある。また、従来のモデルでは、階層的な緊張度を算出する際に、延長的簡約木など、一般的な楽曲に記載されていない情報が必要であった。本論文では、従来のモデルの特徴量に加え、新たな特徴量を導入し、予測モデルを作成する方法について提案する。

2. 関連研究

2.1 ピアノアレンジ

ピアノアレンジの関連研究として、[5][11][12]がある。これらの先行研究は、基本的に原曲スコアに存在する音符から選択してピアノソロ譜を生成している点で共通している。しかし、これらは編曲後の楽曲の雰囲気が維持されていないなど、アレンジの質について考慮しきれていない問題がある。例えば、楽曲中に同じフレーズが何度か出現する際に、1回目は構成音の数が少なく、2回目は1回目比べて構成音の数が多いとすると、構成音の数に応じて編曲後の譜面を変化させるべきである。この問題に対し、高森ら [13] [14] は、原曲のメロディ、コード、リズムに加え

て、厚みといった特徴量をもとにアレンジ譜面を生成した。例えば同じフレーズであっても、同時に鳴っている音の数によって異なる印象を受けると考え、生成される譜面にもその変化が現れるようにしている。この関連研究では厚みという特徴量に着目し、アレンジの質の向上を図っているが、実際に考慮すべき特徴量はそれ以外にも存在すると考えられる。音楽的緊張の観点から述べると、編曲を行う際には、全体の緊張感に関わる重要な特徴量が保たれるように注意する必要がある。

2.2 音楽的緊張のモデル

音楽的緊張に影響を及ぼす特徴量については、[8][6][9][10]や[2][7]などの先行研究がある。[8][6][9][10]の先行研究は、個々の特徴量が緊張感に及ぼす影響について書かれているが、実際の曲はこれらの特徴量の相互作用を受けて全体的な緊張感がある。そこで、各特徴量の相互作用を考えたモデルが必要になる。

Farbood[3][4]が用いた特徴量は相互作用し全体的な緊張感を作り出している。Farboodは音楽的緊張に関与しているとされる各特徴量を、時間経過に従って変化させたいいくつかの例を作成して被験者に聴かせ、その例によって大勢の人が感じる音楽的緊張の外形を調べた、この実験によって明らかになった重要な特徴量をもとに、音楽的緊張のモデルを作成した。このモデルをピアノ弾き語りアレンジの譜面の生成過程に応用する上で問題がいくつかある。研究用に作られた楽曲を用いれば音楽的緊張を予測することができる。しかし、一般的な楽曲は楽譜のみで予めわかっている情報が少ない。例えば、Lerdahl[8][9]の理論をもとに算出するためには、楽曲の延長的簡約木や、部分的な転調が起きている部分など、楽譜に記載されている以上の情報が必要になる。音量による緊張感を正確に算出するには、音響データが必要である。このような理由により、音楽的緊張のモデルを一般的な楽曲に使用することは容易ではない。上記の特徴量の他に、音楽的緊張の予測に貢献する新たな特徴量を探索する必要がある。

3. 音楽的緊張モデルの作成方法

音楽的緊張を以下の式で定義する。各特徴量はその重要度によって重みが掛けられ、線形式によって、音楽的緊張を最も簡易的に表すことができる。

$$y = a_1x_1 + a_2x_2 + \dots + a_nx_n + b$$

- y : 全体的な音楽的緊張
- a_n : 各特徴量の重み
- x_n : 各特徴量の値
- b : 誤差

特徴量 x_n の候補と、その算出方法を 3.1 節に記載する。各特徴量の重みは、教師データと各特徴量の値から重回帰

表 1 特徴量の一覧表

	特徴量の種類
従来研究の特徴量	音高
	音量
	テンポ
	ローカルテンポ
	メロディ期待値
本研究で提案する特徴量	和声
	音の密度
	旋律の傾き
	旋律の平均高さ
	旋律の最高音
	旋律の最低音
非和声音の数	



図 1 音高が上昇したあとに下降する例

分析を行うことで得ることができる。教師データの作成方法は、3.2 節に記載する。

3.1 特徴量の候補と算出方法

本研究に用いる特徴量の一覧を表 1 に示す。本研究では、従来の研究で使用されている特徴量に加え、新たな特徴量を 6 つ提案する。

音高

メロディは、音楽の基本的な構成要素の一つである。曲を聴いたときに、人は音の高さに比例して緊張感を認知している。そこで、音高が半音上がるごとに緊張度が 1 つ増加するように値を設定した。図 1 の楽譜を例にすると、1 小節目の C の音から 2 小節目の 1 オクターブ上の C の音まで、48, 50, 52, 53, 55, 57, 59, 60 と音高の値が上昇し、同じように後半の 2 小節目で下降していく。

メロディ期待値

人は、ある音を聴いたときに次に聴こえてくる音を予測しているとされる。そこで、ある音から次の音へどれだけ遷移しやすいかという、メロディ期待値を特徴量として抽出する。メロディ期待値は、Lerdahl の Melodic attraction (Lerdahl, 2004, pp.161-173) より算出する。(以下の式は参考文献より抜粋)

$$\alpha(p_1 \rightarrow p_2) = \frac{s_2}{s_1} \times \frac{1}{n^2}$$

$\alpha(p_1 \rightarrow p_2)$: ある音 p_1 から次の音 p_2 へのアトラクション

s_i : p_i の anchoring strength (図 2 中の () の中の数値)

n : p_1, p_2 間の半音数

例えば、ハ長調の B から C へのメロディ期待値は $\alpha(B \rightarrow C) = \frac{4}{2} \times \frac{1}{1^2} = 2$ となり、D から C へのメロディ期待値は $\alpha(D \rightarrow C) = \frac{4}{2} \times \frac{1}{2^2} = 0.5$ となる。

(4)	C				G			C	
(3)	C			E	G			C	
(2)	C	D		E F	G	A	B	C	
(1)	C	D♭/C#	D	E♭/D#	E F	G♭/F#	G A♭/G#	A B♭/A#	B C

図 2 ハ長調のメロディのアトラクタ

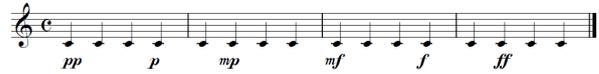


図 3 音量が時間経過とともに増加する例

表 2 音量記号の値

音量記号	値
ff	112
f	98
mf	83
mp	69
p	54
pp	40



図 4 ローカルテンポが時間とともに増加する例

音量

Farbood[3][4]の研究では、音量に比例して緊張度も上昇することが示されている。そこで、表 2 のようにそれぞれの音量記号に対して値を設定した。この値は MusicXML 上の値を使用しているが、適当な値でもよい。図 3 は音量が徐々に増加する例である。この例では、右上がりの緊張感を認知する。

テンポ

テンポは、楽曲の演奏する速さを表している。単位は BPM を用いる。BPM=120 であれば、1 分間の拍数が 120 回入る速さという意味である。緊張感は、テンポにも影響を受けているとされ、テンポが上昇すれば、比例して緊張感が増すとされる [3][4]。

ローカルテンポ

オンセットとオンセットの間隔(秒)は、IOI(inter onset interval)と呼ばれ、音の長さから算出することができる ($IOI = duration \times \frac{60}{tempo}$)。duration は音の長さが何拍分であるかを表す。ローカルテンポは、IOI の逆数を指す。関連研究では、このローカルテンポの値が増えると緊張度が高くなることが示されている [3][4]。図 4 は、曲の進行とともにローカルテンポが増加する例である。

和声

従来の研究では和声の緊張度を算出する際に、Hierarchical tension rule (Lerdahl, 2004, pp.142-161) を用いている。しかし、このルールを用いて算出するには、その楽曲の延長的簡約木を用いて、階層的な緊張度を考慮する必

表 3 ダイアトニックコードと設定した値

コード番号	安定度	値
I	T	0
ii	SD'	2
iii	T'	1
IV	SD	3
V	D	5
vi	T'	1
vii	D'	4

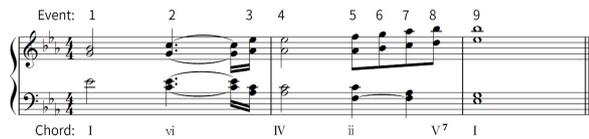


図 5 楽譜の例

要がある。研究用に作成された楽譜の中には、延長的簡約木や、それをもとに算出された和声の緊張度の値が明示されているものもあるが、一般的な楽曲では和声の緊張度を正しく算出することは困難である。そこで、本研究では Lerdahl のルールをもとに和声の緊張度を算出するのではなく、以下に記載する方法で和声の緊張度を簡易的に設定する。

和声による緊張感を簡易的に表すために、そのコードが持つ安定性に従って表 3 のように値を設定した。表中のコード番号は、ダイアトニックコードの番号を表している。各コードは、トニック (T)、トニックの代理コード (T'), サブドミナントの代理コード (SD'), サブドミナント (SD)、ドミナントの代理コード (D'), ドミナント (D) の順番に安定しているとし、0 から 5 までの値を設定した。図 5 の楽譜を例にすると、イベントに対して値が 0, 1, 3, 2, 5, 0 となる。

新たに追加した特徴量

上記の特徴量以外に全体の緊張感を予測するために必要と思われる特徴量を以下に列挙する。

- 音の密度
- 旋律の傾き
- 旋律の高さの平均
- 旋律の最高音
- 旋律の最低音
- 非和声音の数

音の密度は、1 小節中の音数である。音の密度が大きいほど緊張感が高いと仮定する。例えば、図 7 の 1 小節目の音の密度は 5, 2 小節目は 18 である。旋律の傾きは、小節中の最後の音の高さから最初の音の高さを引き、1 小節の時間で割ったものとする。BPM=170 の 1 小節の時間は $4 \times \frac{60}{170} \approx 1.41$ なので、図 7 の 2 小節目の旋律の傾きは、 $(69 - 65)/1.41 \approx 2.84$ となる。旋律の高さの平均は、小節中の旋律のみに着目したときの平均の高さである。図 7 の

1 小節目は、 $(57 + 53 + 48 + 60 + 57)/5 = 55$, 2 小節目は、 $(65 + 64 + 65 + 67 + 65 + 67 + 69) = 66$ と算出される。同じように、図 7 の 1 小節目の旋律の最高音は 60, 最低音は 48 となる。非和声音とは、そのときに鳴っている和音の構成音以外の音のことである。非和声音が多いと緊張感が増すとされる。図 5 を 5 小節目例にすると、Gm7 から C へとコードが進行しているが、Gm7 のコードが鳴っているときは非和声音の数が 4, C のコードが鳴っているときは非和声音の数が 9 となる。

特徴量データの整形

音高、メロディ期待値、音量、テンポ、ローカルテンポの値は線形補間され、それぞれ 50Hz サンプリングにリサンプリングを行なった。これらの値は、MusicXML 形式の楽譜からプログラムによって自動的に値を抽出された。抽出したデータは離散的であり、データの個数が統一されていないため、取得した時系列データの間を補間し、サンプリング周期を統一する必要がある。その他の特徴量に関しては、手動でデータの作成を行った。小節ごと、または半小節ごとに値を設定し、こちらも 50Hz サンプリングに統一した。抽出した全ての特徴量の値は、異なるスケールであったため、それぞれに対して標準化を行った。標準化は特徴量の各データから平均を引き、その値を標準偏差で割ることで達成された。

3.2 教師データの作成方法

被験者実験

聴者が認知する緊張感のデータを収集するために実験を行う。実験方法で注意すべき点は、音楽的緊張に対する聴者のリアルタイムな応答を収集することである。被験者が曲を聴き終わった後に尋ねる方法だと、聴き終わった直前の緊張感の大まかな外形しか反映されないため、リアルタイムで連続的に音楽的緊張への応答を収集する工夫が必要である。そこで、本研究ではテンションスライダーと呼ばれる道具を用いる。図 6 は作成した実験用 WEB サイトである。WEB サイトの画面右にあるのがテンションスライダーであり、値は 0 から 100 まで、初期値は 20 に設定した。被験者は音楽を聴きながら、そのときに認知した緊張度に応じてスライダーを上下に動かす。スライダーが動かされると、そのときの再生時間とスライダーの値が記録される仕組みとなっており、リアルタイムな応答を収集することができる。なお、スライダーの操作性を考え、被験者がスライダーの上にマウスカーソルを置いたときにスライダーのツマミがマウスに追従するように設定した。マウスボタンを押すことなく上下に動かすだけでスライダーを操作できるため、操作ミスが起りにくくなる。WEB サイト画面左にあるのは、再生ボタンと評価値の入力ボタンである。被験者が再生ボタンを押すと、5 秒間カウントダウンした後に曲が流れ始める。カウントダウン機能は、被験



図 6 実験に使用した WEB サイト

者が再生ボタンを押した後に、スライダの位置にマウスを移動し、曲の始まりとともにスライダの操作ができるように実装した。被験者は楽曲を聴き終わった際に、1 試行ごとに評価値を入力する。評価値は、音楽的緊張への応答の信頼度を 5 段階で表している。

実験は、音楽経験のある男女計 3 名対して 3 試行ずつ行った。被験者は、音楽経験の有無や性別に関係なく無作為に選ばれた。実験に使用する音源は、既存の楽曲（ようこそジャパリパークへ）をもとに楽曲作成ソフトで作成されたものである。実験に使用した音源の長さは約 3 分 40 秒であり、被験者 1 人に対し 15~20 分の所要時間を必要とした。被験者に WEB サイトの画面を表示して実験の概要を説明を行ったあと、被験者のタイミングで再生ボタンを押させ、実験を行なった。

実験データの整形

被験者実験によって収集した全ての実験データから、教師データを作成する方法について述べる。収集した実験データは、スライダを動かした際に再生時間とスライダの値を記録したものであった。そのため、時系列データの間を線形補間し、50Hz サンプリグに統一した。その後、それぞれの実験データのスライダの値に対して標準化を行い、データを整形した。被験者ごとにスライダを動かす幅には個人差があったため、標準化を行い、スケールを揃える必要があった。

4. 予測モデルの出力結果と考察

図 7 の 5 小節目までの 1 フレーズの各特徴量の値と正解データを使用し、音楽的緊張の予測モデルを作成した。従来の特徴量 (図 9) と本研究で提案する特徴量 (図 10) を観察すると、ほとんどの特徴量は独立に変化しており、音楽的緊張の予測値を正解データに近づけるために貢献すると思われる。図 9 を見ると、テンポと音量のグラフが重なっていることがわかる。楽曲全体のテンポと音量の値は独立して変化しているが、使用したフレーズにおいては差異がない。1 フレーズという短さでは、テンポも音量も頻



図 7 予測モデルの作成に使用した楽譜

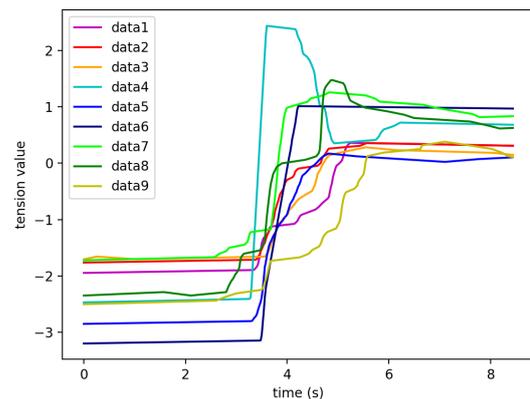


図 8 実験データ

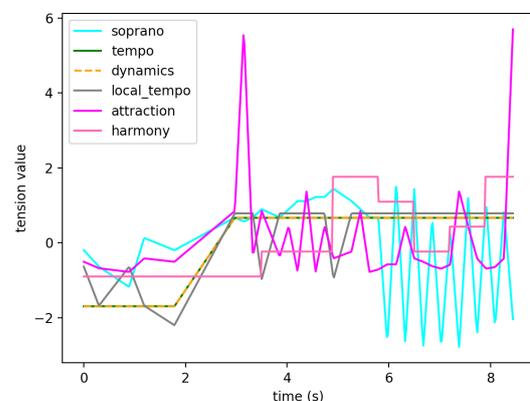


図 9 各特徴量のグラフ

繁に変動しないため、細かな緊張度の予測には貢献度が低いと考えられる。図 11 の青色のグラフは教師データの値で、赤色は予測値である。教師データは、図 8 に示してある、標準化した実験データの平均とした。予測値は全体的に正解データに近似されていることがわかる。しかし、個々の特徴量の影響を受け、正解データとの大きな差がある箇所も存在する。

5. おわりに

本研究は、ピアノ弾き語り譜の生成に応用可能な音楽的緊張の予測モデルの構築を目指す。従来の研究では、研究

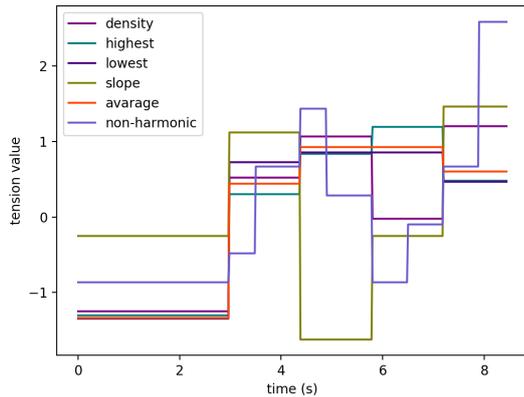


図 10 各特徴量のグラフ

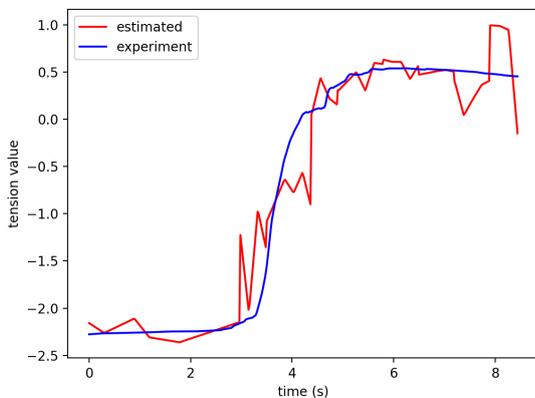


図 11 音楽的緊張の予測モデル

用に作成されたクラシック楽曲を対象としているため、ピアノ弾き語り譜の生成に应用するためには、特徴量を見直す必要があった。そこで、本論文では、従来の特徴量に新たな特徴量を加えて予測モデルを構築する方法について提案を行った。今後は予測モデルの評価を行う。

謝辞 本研究に取り組むにあたり、ご助言をくださった寺井あすか准教授（公立はこだて未来大学）に心より感謝いたします。本研究は JSPS 科研費 16H01744 の助成を受けたものです。

参考文献

- [1] Ball, Philip.: The Music Instinct: How Music Works and Why We Can't Do Without It. London, Bodley Head (2010). 夏目大 (訳) : 音楽の科学: 音楽の何に魅せられるのか?, 河出書房新社.
- [2] Bigand, E., Parncutt, R., & Lerdahl, F.: Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training, *Perception & Psychophysics*, Vol.58, No.1, pp.125-141 (1999).
- [3] Farbood, M. M.: A Quantitative, Parametric Model of Musical Tension, Ph.D. Dissertation, Massachusetts Institute of Technology (2006).
- [4] Farbood, M. M.: A global model of musical tension,

- Proc. the 10th International Conference of Music Perception and Cognition, pp.690-695 (2008).
- [5] 藤田顕次, 大野博之, 稲積宏誠: 習熟度を考慮した複数楽譜からのピアノ譜生成手法の提案, 情報処理学会研究報告音楽情報科学 (MUS), Vol.2008-MUS-077, No. 89, pp.47-52 (2008)
 - [6] Jehan, T.: Creating music by listening. Diss. Massachusetts Institute of Technology, School of Architecture and Planning, Program in Media Arts and Sciences (2005).
 - [7] Krumhansl, C. L.: A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas, *Music Perception: An Interdisciplinary Journal*, Vol.13, No.3, pp.401-432 (1996).
 - [8] Lerdahl, F.: Tonal pitch space, Oxford University Press (2004).
 - [9] Lerdahl, F., & Krumhansl, C. L.: Modeling tonal tension, *Music Perception, An Interdisciplinary Journal*, Vol. 24, No. 4, pp.329-366. (2007).
 - [10] Margulis, E. H.: A model of melodic expectation, *Music Perception: An Interdisciplinary Journal*, Vol.22, No.4, pp.663-714 (2005).
 - [11] Nakamura, Eita, and Shigeki Sagayama.: Automatic Piano Reduction from Ensemble Scores Based on Merged-Output Hidden Markov Model, *ICMC* (2015).
 - [12] Onuma, Sho, and Masatoshi Hamanaka.: Piano Arrangement System Based On Composers' Arrangement Processes, *ICMC* (2010).
 - [13] 高森啓史, 佐藤晴紀, 中塚貴之, 森島繁生: 原曲の楽譜情報に基づいたピアノアレンジ譜面の生成, 情報処理学会第 79 回全国大会講演論文集, Vol.2017, No.1, pp.91-92 (2017) .
 - [14] 高森啓史, 佐藤晴紀, 中塚貴之, 森島繁生: 原曲スコアの音楽特徴量に基づくピアノアレンジ, 研究報告音楽情報科学 (MUS), Vol.2017-MUS-114, No.16, pp.1-6 (2017) .
 - [15] 東条敏, 平田圭二: 音楽・数学・言語: 情報科学が拓く音楽の地平, 近代科学社 (2017).