

歌唱のグルーブ感の構成要因の分析

的場 達矢^{†1} 馬場 隆^{†1} 成山 隆一^{†2} 松本 秀一^{†2} 森勢 将雅^{†3} 片寄 晴弘^{†1}

概要: VOCALOID などの音声合成技術の普及に伴い、歌声に関連する研究は活性化し、歌声情報処理と呼ばれる研究領域が定着した。また、歌声のうまさを自動採点する技術が実用化され、プロ歌手が「うた」の上手さを競う TV 番組が制作されるなど、歌のうまさについて興味が高まっている。Pops 歌唱の主要な表現対象の一つに「グルーブ感」が存在するが、その構成要因については明らかになっていない。本稿では、プロ歌唱者による「グルーブ歌唱」と「非グルーブ歌唱」の比較に基づいて、聴取者が「グルーブ感」を感じる要因が何であるのかについて検討した結果について報告する。分析の結果、子音長が「グルーブ感」の重要な構成要因であることが見いだされた。

1. はじめに

VOCALOID などの音声合成技術の普及に伴い、アマチュアによる音楽制作では、合成歌唱音声を用いた音楽制作活動が積極的に行われるようになってきた。これに対応する形で、歌唱に関連する研究も活性化し [1], [2], 歌声情報処理 [3] と呼ばれる研究領域が定着するに至っている。

「うた」は、動画共有サイト、カラオケ録音公開サイトなどの「歌ってみた」動画の盛り上がりにも見られるように、最も身近な自己表現の対象であり、その上手さは、誰でも判定が可能である。また、下手な歌唱には、ピッチや音量が不安定であったり、ビブラートやしゃくり等の歌唱技法が不安定であるというような特徴があげられる。このような特徴を捉えた自動歌唱採点機能が実用化され、現在のカラオケシステムの人気機能の1つとなっている。最近では、ロングトーンやビブラートによる加点や、ビジュアライゼーションの強化が進み、プロ歌手が「うた」の上手さを競う TV 番組が制作されるなどの盛り上がりを見せている。カラオケ採点システムの採点精度は、登場した当初と比べて大幅に改善したが、レベルの高い歌唱間へのシステムが割り出す採点結果は必ずしもヒトの判定と同じにならないことが広く一般にも認知されている。「グルーブ感」や「感情表現」といった高次の「うた」の上手さを判定する基準がどのようなものなのかについては、現状では解明されていない。

本研究では、Pops 歌唱を対象として「グルーブ感」を構成する要因を分析することを目的とする。十分に高いレベ

ルのプロ歌唱者の「グルーブ歌唱」と「非グルーブ歌唱」の比較に基づいて、聴取者が「グルーブ感」を感じる要因が何であるのかについての検討を実施する。以下、第2章では、本研究のアプローチについて、第3章では、データの収集と分析について述べ、歌唱のグルーブ感の構成要因の仮説について述べる、第4章では、聴覚実験による検証について述べ、第5章では、まとめと今後の展望について述べる。

2. グルーブ感と課題解決のアプローチ

グルーブ感とは、ジャズ、レゲエ、ソウルなどのブラックミュージックの音楽独特の演奏表現様式をルーツとし、主としてリズムやノリに関連した表現様式として理解されている。その具体的な定義は定まっていないが、ドラムスにおいては、スネアドラムの打点位置 [4] や、ゴーストノートが重要な役割を果たしている [5] ことが見いだされている。また、ドラムスやベースなどのリズムセクションにおける表現として語られることが多いが、歌唱においてもグルーブ感を意識した表現が求められることが少なくない [6]。その差の要因を明文化することは難しいが、グルーブ感の表出を意識した歌唱とそうでない歌唱では、ヒトは明らかな差異を感じる事ができる。

近年、歌唱合成ソフトウェアの普及に伴って歌唱データの作り込みが行われるようになり、自然な歌声を生成することを目的として、音素の配置や大きさ、音色の微妙な調整が行われるようになった。この過程を通じて、例えば、「促音*1」の表現について場合分けが必要であること等、普段我々の意識に上ることが少ない「聞こえ」の背景となる

^{†1} 現在、関西学院大学

^{†2} 現在、ヤマハ株式会社

^{†3} 現在、山梨大学

*1 日本語のかな表記において「っ」で表されるつまる音

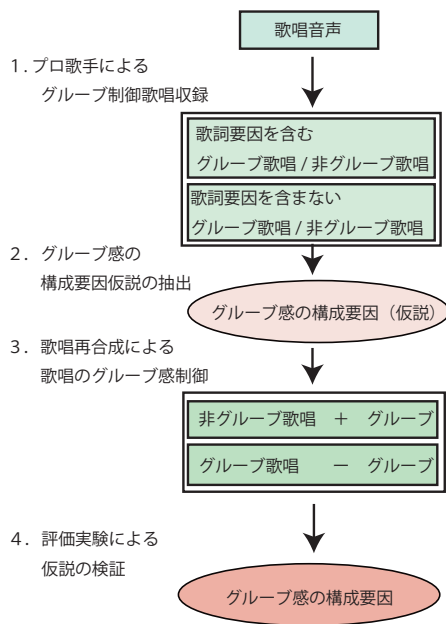


図 1 本研究のアプローチの概要

要因について形式化が進んだものもある。この例が示すように、グループ感の構成要因を明らかにしていく方法論の1つに、analysis by synthesisによる方法（構成論的アプローチ）が考えられるが、作業仮説の確保自体が困難であるという課題がある。ここでは、高レベルで「グループ感」の表現が可能なプロ歌手の協力のもと、「グループ歌唱」と「非グループ歌唱」を収録し、その音響的な比較によって、その差異を構成する要因についての仮説を見いだす。その仮説を基に生成した合成歌唱の聴取実験によって、「グループ歌唱」の構成要因を明らかにしていく。図1に本研究のアプローチの概要を示す。

2.1 分析対象

Pops音楽には、音楽ジャンルなどの異なる様々な楽曲が存在する。その中でも、グループ感の表出しやすい楽曲と表出しにくい楽曲が存在する。ここでは、J-Popsの中でもブラックミュージックに関連の深いアーティストとして知られている久保田利伸の楽曲を取り上げる。具体的には、異なるグループ感の表現様式とされる「LALALA LOVESONG」（以降、楽曲A）、「LOVERAIN～恋の雨～」（以降、楽曲B）を取り上げる。

2.2 「マ行」統制歌唱の利用

ボーカロイドなどの合成歌唱を用いた音楽制作において、/s/などの子音を含む音節は他の子音を含む音節と比較して前にずらされる。この例が示すように、子音の種類が歌唱表現に影響を与えており、歌唱のグループ感の構成要因に影響を与える要因となり得ることが考えられる。そこで、子音変化の要因を統制するものとして、歌詞の子音をすべて/m/に変えた「マ行」統制歌唱を用意し、通常歌

詞歌唱との比較によって、グループ感の構成要因の仮説の獲得を目指す。

2.3 歌唱者

高い歌唱力を有するプロ歌手と一言で言っても、アーティストとして活躍するプロ歌手、主としてコーラスを担当とするスタジオミュージシャン、ボイストレーナーとして歌唱教育に関わる幅広いプロ歌手が存在する。本研究では、J-Popsのプロ歌手として高い歌唱力を持ち、かつ、プロ歌手のボイストレーナーとして歌唱表現の指導に携わっている歌手を協力者として迎えて検討を進める。

3. 歌唱データの収録と分析

本章では、歌唱データの収録と分析の実際について述べる。グループ感は、音響情報を手がかりとして知覚される。音響情報中のさまざまな情報、また、メロディラインや拍節構造、和声などの楽曲固有の要因との相互作用によってグループ感が形成されると考えられるが、ここでは、歌唱のオンセット時刻を中心とした分析を実施する。

3.1 歌唱データの収録

「グループ歌唱」および「非グループ歌唱」の収録に際して、第二章にて述べたJ-Popsのプロ歌手として高い歌唱力を持ち、プロ歌手のボイストレーナーとして歌唱表現の指導に携わっている男性歌手を協力者として迎え、第二章にて述べた分析対象となる歌唱音声収録した。

収録に際して、歌唱者には、「グループ歌唱」として、「通常の歌唱」と依頼し、「非グループ歌唱」として、「グループのない歌唱」、「棒歌い」と依頼した。収録では、レコーディングスタジオにて、一定テンポのカラオケ音声をヘッドフォンにて聴取しながらの歌唱音声を収録した。なお、収録歌唱は、第二章にて述べた楽曲AおよびBにおける、「マ行」統制歌唱および通常歌詞歌唱の「グループ歌唱」および「非グループ歌唱」であり、計8歌唱である。

3.2 分析手法

「グループ感」に関する従来研究においては、主として、ドラムスやピアノを対象に、音のオンセット、オフセット、インテンシティの分析がなされてきた[4],[5]。歌唱においては、これらに加えて、F0、音圧、音色に関する経時的な変化を分析していくべきである。本研究では、歌唱のグループ感分析の第一段階として、子音および母音それぞれのオンセットに着目して分析を進めていくこととする。

子音および母音の区間を自動抽出するためのツールや技術については一部実用化されているものもあるが、これらツールの精度は100%には及ばない。本研究では、市販のソフトウェアを用いて子音相当の部分を手作業で削っていき、試聴した結果、音韻が変化したタイミングを特定していく

表 1 抽出子音および母音の種類一覧

音素	抽出	種類	音素	抽出	種類
/a/	○	母音	/i/	○	母音
/u/	○	母音	/e/	○	母音
/o/	○	母音	/k/	○	子音 (破裂音)
/s/	○	子音 (摩擦音)	/g/	○	子音 (破裂音)
/sh/	○	子音 (摩擦音)	/z/	○	子音 (摩擦音)
/t/	○	子音 (破裂音)	/j/	○	子音 (摩擦音)
/ch/	○	子音 (摩擦音)	/d/	○	子音 (破裂音)
/h/	○	子音 (摩擦音)	/n/	○	子音 (鼻音)
/m/	○	子音 (鼻音)	/b/	○	子音 (摩擦音)
/r/	○	子音 (R 音)	/y/	×	子音
/w/	×	子音	/	/	/

という手段で、ms 単位の分解能でオンセットの分析を行う。また、作業者による結果の差異を吸収するものとして、3名の作業者が個別に分析を実施し、差異が見つかった際には、3名が集まって再分析をするという手続きを採った。

以下に本研究におけるオンセット情報抽出処理の具体的な手順について示す。

- (1) 歌唱音源の聴取および歌唱波形の目視から、抽出対象とする子音オンセットをラフに同定し、直前区間から該当のオンセットまでを再生区間として指定する。
- (2) 指定された再生区間の聴取を行い、抽出対象とする子音オンセットの子音知覚について確認を行い、子音が知覚された場合、再生区間の再指定を行う。
- (3) 1および2を繰り返し行い、抽出対象とする子音の知覚開始時刻を推定し、該当の時刻を子音オンセットとする。

子音のオフセット時刻については、後続の母音、もしくは、子音のオンセット時刻で与えることにした。母音オフセット情報は、残響音の影響により、正確性にかけるため今回の分析からは対象外とした。

本研究では、日本語における子音を、摩擦音、破裂音、鼻音、R音の4種類、母音を、/a/, /i/, /u/, /e/, /o/に分類し、分析を実施した。分析対象を表1に示す。

3.3 分析結果

楽曲 A 「マ行歌唱」と「通常歌詞歌唱」における子音オンセットおよび母音オンセットの譜面上の拍点からのズレのヒストグラムを、それぞれ、図 2, 3 に示す。母音オンセットについては、「マ行歌唱」と「通常歌詞歌唱」の双方において、「グループ歌唱」においても「非グループ歌唱」においてもほぼ同等の分布傾向が認められる。これに対して、子音オンセットについては、「マ行歌唱」において「グループ歌唱」において開始点が早くなるものがあることが確認される。「通常歌詞歌唱」においては、「グループ歌唱」においても「非グループ歌唱」においても子音オンセットは大きく散らばっている。このことは、**母音オンセットがリ**

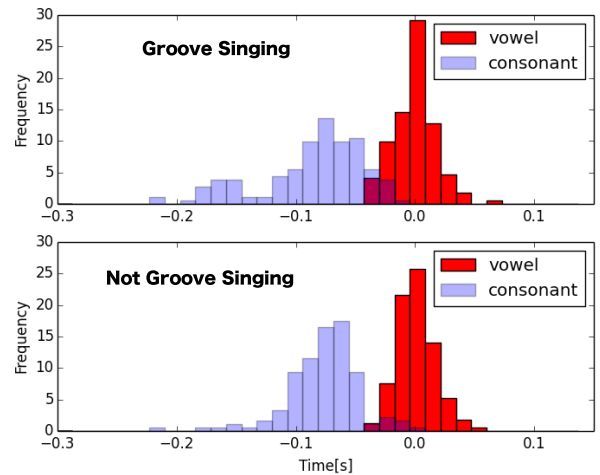


図 2 「m」子音統制歌唱における楽曲 A の子音オンセット及び母音オンセットのズレのヒストグラム

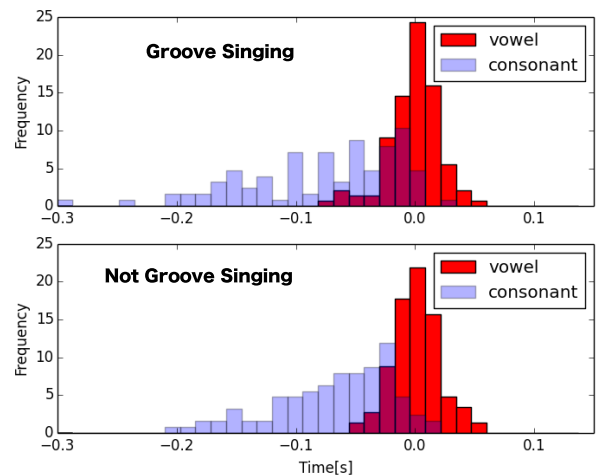


図 3 通常歌詞歌唱における楽曲 A の子音オンセット及び母音オンセットのズレのヒストグラム

ズムの正確性に寄与し、鼻音 (nasal) を含む特定の子音の長さ (子音長) が歌唱のグループ感に関与しているという可能性を示唆している。

楽曲 A における「マ行歌唱」、「通常歌詞歌唱」の子音長のヒストグラムを、それぞれ、図 4, 5 に示す。「マ行歌唱」においては、「非グループ歌唱」では、約 0.08[s] 付近で子音長が分布しているのに対して、「グループ歌唱」では 0.2[s] を超えるあたりまで広がっている。図 5 (通常歌詞歌唱) からは、破裂音 (burst)、摩擦音 (brush)、R 音 (R)、鼻音 (nasal) の中で、鼻音 (nasal)、摩擦音 (brush) の子音長の分布が散らばっている様子が読み取れる。

以上から、**鼻音 (nasal)、摩擦音 (brush) の子音長が歌唱のグループ感の構成要因として関与している**という仮説を得ることができる。

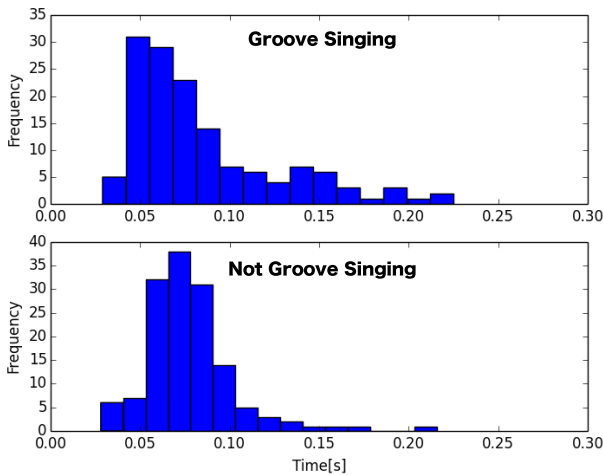


図4 「m」子音統制歌唱における
 楽曲 A の子音長のヒストグラム

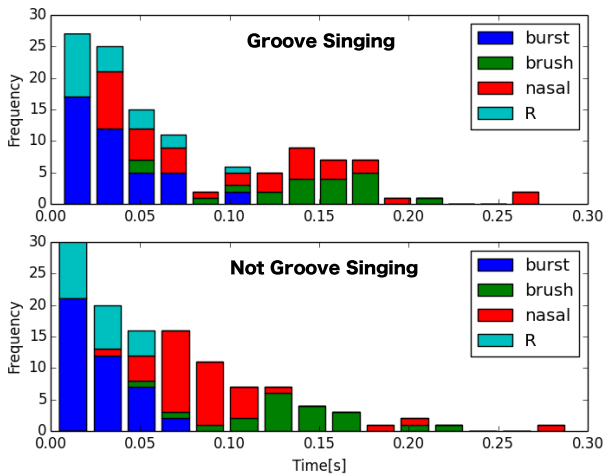


図5 通常歌詞歌唱における
 楽曲 A の子音長のヒストグラム

4. 歌唱のグルーブ感の聴取評価実験

第三章で獲得した歌唱のグルーブ感の構成要因仮説である母音オンセットおよび子音長の歌唱のグルーブ感への影響の評価を目的として、各種特徴量を変化させた歌唱音声を作成し、聴取実験を実施する。

4.1 実験刺激

実験刺激は、子音長および母音オンセットがグルーブ感評価に与える影響を調査するために、合成元音声の子音長、母音オンセット、もしくはその両方を「グルーブ歌唱」「非グルーブ歌唱」の情報へと変化させたものを用いる。また、グルーブ感の構成要因には、三章にて取り上げた歌唱の時間変化する特徴量が含まれていることが考えられ、それらの影響により、母音オンセットおよび子音長の

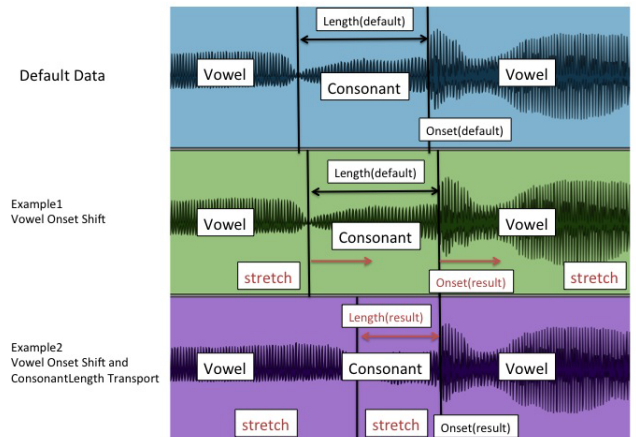


図6 処理の概要

変化がグルーブ感評価へ与える影響が変化することが考えられる。そのため、歌唱の時間変化する特徴量として、第三章にて収録した「グルーブ歌唱」、「非グルーブ歌唱」を合成元音声として利用することで、歌唱の時間変化する特徴量を包含し、2つの事例として用いる。合成歌唱データの生成では、TANDEM-STARIGHT[7]におけるF0推定とWORLD[8]によるスペクトル解析および、再合成を用いた。図6に子音長および母音オンセットの制御処理の概要を示す。これにより、表2に一覧を示す合成歌唱データを生成し、実験刺激として用いる。

グルーブ感の構成要因に影響を与える要因は、メロディラインや拍節構造、和声、歌詞などが考えられる。本研究では、グルーブ感の構成要因に影響を与える要因としてメロディラインの要因をとりあげ、異なるメロディラインを伴う複数の歌唱区間を選択する。メロディラインの選択では、収録楽曲における歌唱音声より、Eugene Narmourにより考案されたImplication-Realization Model[9]を利用し、音高の跳躍の大きさを考慮した3つの異なるメロディラインの歌唱区間を選択し、刺激として用いる。具体的には、Implication-Realization Modelにおける「P/P/P/IP」(以降、歌唱区間A)、「ID/VR/VR/IR」(以降、歌唱区間B)、「P/D/ID/ /P/D/ID」(以降、歌唱区間C)の3区間である。歌唱区間AおよびBは楽曲Aより選択し、歌唱区間Cは楽曲Bより選択した。選択された歌唱音声のそれぞれの長さは約4秒であった。

4.2 実験方法

聴取実験は、シェッフエの対比較法(中屋の変法)[10]により行った。それぞれの歌唱区間における8種の刺激音のうちいずれかの2つを対にして呈示し、「どちらがグルーブを感じるか」について、7段階の評価尺度(-3~3)で評価させた。各歌唱区間における刺激対の総数は28組となり刺激の総数は84組となる。また刺激対は順序効果を考

表 2 1つの歌唱区間における刺激となる歌唱データの一覧

刺激 ID.	合成元音声	子音長	母音オンセット
A	グループ歌唱	グループ歌唱	グループ歌唱
B	グループ歌唱	グループ歌唱	非グループ歌唱
C	グループ歌唱	非グループ歌唱	グループ歌唱
D	グループ歌唱	非グループ歌唱	非グループ歌唱
E	非グループ歌唱	グループ歌唱	グループ歌唱
F	非グループ歌唱	グループ歌唱	非グループ歌唱
G	非グループ歌唱	非グループ歌唱	グループ歌唱
H	非グループ歌唱	非グループ歌唱	非グループ歌唱

表 3 歌唱のグループ感に関する相対的嗜好度

刺激 ID.	歌唱区間 A	歌唱区間 B	歌唱区間 C
A	0.7	4.12	2.9
B	0.8333	8	-3.5
C	-5.2583	-0.205	-0.35
D	-4.3333	3.325	-0.5
E	-36.0083	-32.18	-34.85
F	-48.6833	-31.46	-36.325
G	-40.1083	-39.365	-38.05
H	-43.4083	-36.395	-38.525

慮してランダムに呈示した。被験者は、37名の成人(男性28名,女性9名)である。被験者は各自の自宅でWeb上に呈示された実験システムを用いて、実験を行った。被験者は、実験の回答時間に制限はなく、刺激音を一回以上聴取した後、その刺激音に対して評価を行った。被験者は刺激音の聞き直しを無制限に許可した。被験者は、上記の評価実験終了後、ランダムに呈示された刺激対「A,B」「A,C」「F,H」「G,H」に対し、その違いについて自由記述にてコメントした。

4.3 実験結果

表3に、各歌唱区間における歌唱のグループ感に関する相対的嗜好度の一覧を示す。

各歌唱区間に対して、最もグループ感があると判断された刺激は全て、合成元音声および子音長が「グループ歌唱」の事例であった。また、各歌唱区間において、合成元音声が「非グループ歌唱」の事例の刺激の中で、最もグループ感があると判断された刺激は、子音長が「グループ歌唱」の刺激であった。以上のことから、子音長は、歌唱のグループ感に関する要因であることが示唆された。

被験者が子音長が歌唱のグループ感に関する要因であることを支持されていた。

刺激対における、子音長が「グループ歌唱」の事例に対して、「グループ感がある」「抑揚がある」「強弱の変化がある」「表情豊かに聞こえる」などのコメントが得られた。子音長が「非グループ歌唱」の事例に対して、「平坦に聞こえる」「棒読み聞こえる」などのコメントが得られた。

5. まとめと展望

本研究では、プロ歌唱者に協力を仰ぎ、/m/子音を用いた子音統制歌唱、通常歌詞を用いた子音混合歌唱における、「グループ歌唱」および「非グループ歌唱」を収録した。収録した「グループ歌唱」および「非グループ歌唱」に対して、歌唱のオンセット時刻を中心とした分析を実施した。「グループ歌唱」および「非グループ歌唱」における子音オンセット、母音オンセット、子音長の分析結果、子音長が歌唱のグループ感の構成要因として関与している仮説を立てた。次に、聴取実験を通して、子音長がグループ感評価に寄与するかを調査した。実験では、WORLDを用いた歌唱再合成により、子音長の制御を行い、そのグループ感変化を調査した。その結果、子音長が、歌唱のグループ感変化に影響を与えることが確認された。

今後は、本研究にて収録した「グループ歌唱」「非グループ歌唱」に対して、歌唱音声におけるF0、音圧、音色の経時的な変化の分析を進める。また同時に、和声や拍節構造、歌詞、メロディラインなどを考慮した「グループ歌唱」「非グループ歌唱」を収録・分析することで、歌唱におけるグループ感や感情表現における子音長の制御要因の解明を目指す。

参考文献

- [1] 中野 倫靖, 後藤 真孝, VocaListener: ユーザ歌唱の音高および音量を真似る歌声合成システム, 情報処理学会論文誌, Vol.52, No.12, pp.3853-3867, December 2011.
- [2] 辰巳 直也, 馬場 隆, 森勢 将雅, 片寄 晴弘: ロックボーカルレゾネータ VocaLoid 歌唱をロックボーカリスト風の歌い方に変換するシステム, 研究報告音声言語情報処理 (SLP), 2012.
- [3] 後藤 真孝, 斎藤 毅, 中野 倫靖, 藤原 政将: 歌声情報処理の最近の研究, 音響学会論文誌, Vol.64, No.64, No.10, pp.616-623 2008.
- [4] 奥平啓太, 平田圭二, 片寄晴弘: ポップス系ドラム演奏の打点時刻及び音量とグループ感の関連について (第2報), 情報処理学会研究報告, Vol.14, pp.27-32, 2005.
- [5] 奥平啓太, 平田圭二, 片寄晴弘: ポップス系ドラム演奏の打点時刻及び音量とグループ感の関連について (第3報), 情報処理学会研究報告, Vol.2004-MUS-084, pp.53-58, 2006.
- [6] 鈴木泰司: DVD でよくわかる もう一度聴きたい! と言わせるボーカルテクニック, 日東書院, 2013.12.
- [7] Hideki Kawahara, Toshio Irino and Masanori Morise, An interference-free representation of instantaneous frequency of periodic signals and its application to F0 extraction, Proc. ICASSP 2011, pp.5420-5423, May 2011.
- [8] 森勢将雅, 西浦敬信, 河原英紀, 「高品質音声分析変換合成システム WORLD の提案と基礎的評価-基本周波数・スペクトル包絡制御が品質の知覚に与える影響」日本音響学会聴覚研究会, Vol. 41, No. 7, pp. 555-560, Toyama, Oct. 2011.
- [9] Eugene Narmour: "The Analysis and Cognition of Melodic Complexity" The university of Chicago press, 1992.
- [10] 佐藤 信: 統計的官能検査法, 日科技連, 1985.