

聴覚フィードバックを用いた歌唱時の音程操作

井手口 裕太^{1,a)} 横畑 亮輔¹ 井村 誠孝¹ 浦西 友樹¹ 黒田 嘉宏¹ 大城 理¹

概要: 近年、カラオケなどで歌唱する機会が増加し、歌唱力を向上させるシステムの開発が望まれている。本研究では、歌唱者の音程のずれに応じた伴奏音を提示することにより、歌唱者の音程を操作し、歌唱力を向上させるシステムの開発を目指す。歌唱者の音程のずれに応じた伴奏音を提示するため、歌唱者の音程のずれの特徴を定量化する必要がある。本稿では、特徴を定量化するため、歌唱者に任意の伴奏音を提示した際の音程のずれを実験的に調べた結果を示す。

1. はじめに

人は古来より多くの場面で歌を歌い、現在ではカラオケの普及などにより一般の人が人前で歌う機会が急増している。特にカラオケなどでは、歌唱力はコミュニケーションに影響を与える要因であり、歌唱力向上を目指す人が多くおり、歌唱力を向上することが可能なシステムが求められている。人は歌唱する際に、伴奏音や周囲の環境音などに合わせて歌唱する。これにより、伴奏音が違っていたり複数人で歌唱する際に、歌唱者の音程がこれらの環境音に「つられる」といった現象が発生する。本研究では、この「つられる」という現象を利用し、環境音を適切に操作することで歌唱者の発する歌声の音程を意図的に操作するシステムを構築することを目指す。本システムの利用により、原曲の音程と同じ音程で歌うことが可能となる。

歌唱音の音程を操作する手法は、出力音程を直接処理する手法と、聴覚的な指導または視覚的な指導により出力音程を変化させる手法の二つに分けることができる。しかし、言語的、視覚的な矯正は、歌唱者の意識的な努力が必要であり、歌唱する楽しみを阻害する可能性がある。本システムは、図1に示すように歌唱者ごとの環境音に対する音程のずれを学習し、ずれを考慮して音程をわずかに変調させた伴奏音を提示することによって、歌唱者の音程を操作する。これにより、歌唱者に操作されている感覚を与えずに、自然に伴奏音を聞きながら歌唱することで、結果的に音程を操作することが可能であると考えられる。本システムは、歌唱者にはヘッドホンにより個人の音程のずれに合わせて伴奏音を提示し、周囲の人には原曲と歌唱者の歌声を提示する。歌唱者個人の音程のずれに応じた伴奏音を

提示するためには、歌唱者個人の音程のずれの特徴を定量化する必要がある。本稿では、特徴を定量化するため、歌唱者に任意の伴奏音を提示した際の音程のずれを実験的に調べた結果を示し、伴奏音の変調による歌唱音の操作の可能性について検証する。

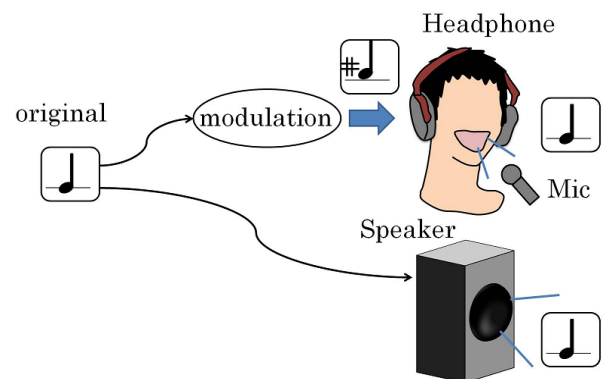


図1 システム概念図

2. 歌唱する際の音程制御

人は歌唱する際に、周囲の様々な要因に影響されながら歌唱している。目的の音程を発声するための手がかりは、感覚性のもものと、発声する際の運動性のもとの二つに大きくわけることができる。感覚性は、音程に対する認知・理解・記憶であり、運動性は発声する際の舌の動きや筋肉の動きによるものである [1]。また、人は伴奏音を聞きながら歌唱する際、自分の歌声を聴きフィードバックを掛けることで、音声を伴奏音に近づくように歌唱しており [2][3]、歌唱者の音程を視覚的にフィードバックさせることで、出力の音声を変化することが可能であることが示されている [4][5][6]。また、環境音に対する歌唱者の音声の分析やモデル化といった研究なども行われている [7][8][9]。これ

¹ 大阪大学 大学院基礎工学研究科
Graduate School of Engineering Science, Osaka University
^{a)} ideguchi@bpe.es.osaka-u.ac.jp

らのことから本研究では、環境音に対する歌唱者の音程をモデル化し、環境音を適切に操作することで歌唱者の音程を意図的に操作するシステムを構築する。

3. 提案手法

3.1 概要

本システムは、歌唱者の環境音に対する音程のずれを学習し、ずれを考慮した伴奏音を提示することによって、歌唱者の音程を操作する。歌唱音の音程は、環境音を構成する様々な要因に基づいて歌唱者が決定すると考えられるため、環境音を構成する複数の要因 f_{in} と人の音声の音程 f_{out} の関係は式 (1) に示されるように形式的に表される。

$$f_{out} = G(f_{in}) \quad (1)$$

システム概要を図 2 に示す。本システムは、環境音に対する歌唱者個人の音程のずれの特徴を定量化するキャリブレーションと、キャリブレーションにより求めた特徴を考慮した伴奏音の提示の二つに分けることができる。歌唱者はヘッドホンにより、周囲の人とは異なる伴奏音を聞かせることで、周囲の人は原曲を聴きながら歌唱者の歌声を聴くことができる。

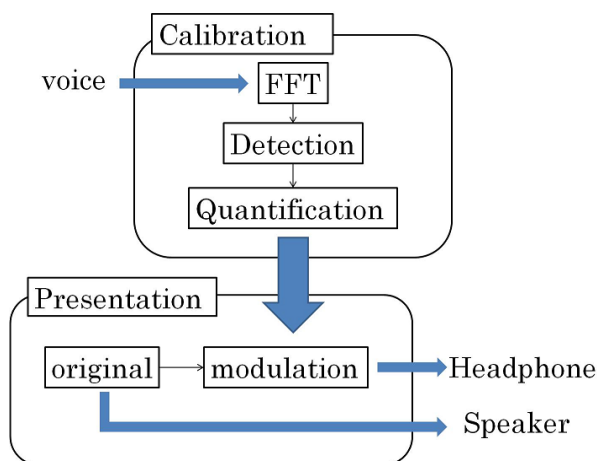


図 2 システムの概要

3.2 キャリブレーション

本研究では、歌唱者の音程のずれは常に一定のずれを示すと仮定する。伴奏音に対する歌唱者の音程のずれを調べ、歌唱者個人の特徴を定量化する。音程を調べるために、音声信号をフーリエ変換し、音程を検出する。

歌唱者個人の特徴として以下の項目が考えられる。

- 単音の伴奏音を提示した際の音程のずれ
- 連続音の伴奏音を提示した際の音程のずれ
- 伴奏音に対する遅れ

本稿では、単音の伴奏音のみに着目し、伴奏音の音程 f_{in} と人の音程 f_{out} との関係は以下のように示すことができる

ものとする。

$$f_{out} = G(f_{in}) \quad (2)$$

関数 G が歌唱者の伴奏音に対する音程のずれ方を表すモデルであり、一般的には f_{in} に対する単調増加関数であると考えられる。

3.3 伴奏音の提示

キャリブレーションにより定量化した個人の特徴を基に操作したい音程に合わせて伴奏音を提示する。歌唱者に歌わせたい音程を f_{out}^* とすると、提示する伴奏音 f_{in}^* は式 (3) により、求めることが可能となる。

$$f_{in}^* = G^{-1}(f_{out}^*) \quad (3)$$

伴奏音の提示は、歌唱者にヘッドホンから伴奏音を提示することで、歌唱者のみに伴奏音を提示する。伴奏音を提示する際には以下のことに注意する必要がある。

- 歌唱者に操作していることを感じさせない。
- 原曲に比べて、不自然な音程の変化をさせない。

4. 実験

4.1 手法

本システムは、歌唱者個人の音程のずれに応じた伴奏音を提示するため、歌唱者個人の音程のずれの特徴を定量化する必要がある。特徴を定量化するため、歌唱者に任意の伴奏音を提示した際の歌唱者の音程のずれを実験的に調べるため被験者実験を行った。

今回は、単音の伴奏音を提示した際の音程のずれを調べるため、ある音程の伴奏音を 2 秒間流し、被験者はその音をヘッドホン (SE-MJ521 Pioneer 製) で聴きながら「aa」と発声してもらい、マイク (ECM-PCV80U SONY 製) を通してサンプリング周波数 44.1kHz の 16bit のモノラル信号として録音し、その時の音程を調べた。音程は録音した音声を矩形窓関数で短時間フーリエ変換し、ピークとなる周波数を音程とした。音声は、実際の音程の周波数の整数倍ごとにピークが現れることになる。従って、最も低い周波数のピークを検出することで、録音した音声の音程を求めることができる。ピーク検出には、まずフーリエ変換した波形の一次近似直線成分を減算する。その後移動平均し、閾値 α でフィルタリングする。 α は最大値の 0.65 倍とした。この信号に対し周波数の低い方から極大となる点を検出し、この近辺の最大となる点を検出することでピーク検出した。

被験者実験は、実験に使用する音程すべてを認知することが可能な 22~23 歳の健常男性 6 名で行った。提示した伴奏音は男性が発声するのに無理のない音程の伴奏音を提示した。実験には MIDI のアコースティックピアノの音源を使用し、表 1 に示すノートナンバー 48~60 (130.8~

表 1 ノートナンバー

ノートナンバー	音名	周波数 [Hz]
48	C ド	130.8
49	C # ド#	138.6
50	D レ	146.8
51	D # レ#	155.6
52	E ミ	164.8
53	F ファ	174.6
54	F # ファ#	185.0
55	G ソ	196.0
56	G # ソ#	207.7
57	A ラ	220.0
58	A # ラ#	233.1
59	B シ	246.9
60	C ド	261.6

261.6Hz) の伴奏音をランダムに約 400 回提示した。

この実験は、大阪大学大学院基礎工学研究科 人を対象とした研究に関する倫理委員会の承認 (24-6) を得ており、被験者の同意・協力のもと行ったものである。

4.2 結果

被験者に提示した伴奏音 (ド: 130.8Hz) の例及び、その際の被験者の音声の例を図 3、4 に示す。これをフーリエ変換したものを図 5、6 に示す。音声の音程は図 7 に示すように、時間により変動するため、音程が安定した区間を平均し、その音声の音程とした。各伴奏音ごとの音程を、ある被験者についてまとめたものを図 8 に示す。他の被験者の実験結果を図 9~13 に示す。また、図 8~13 の黒線は各周波数での音程の平均値を示している。

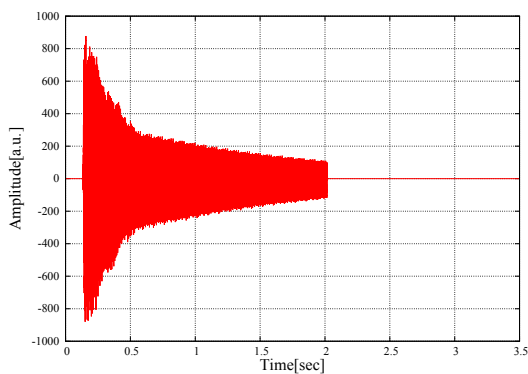


図 3 被験者に提示した伴奏音 (ド: 130.8Hz)

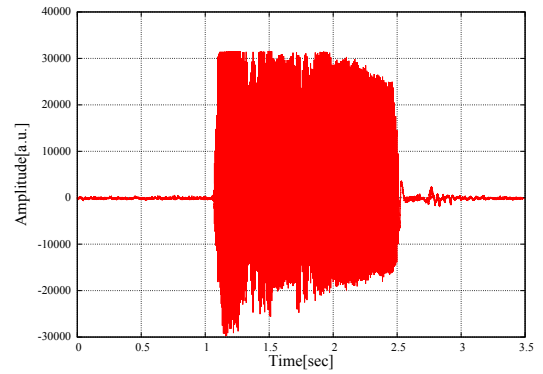


図 4 被験者 A の音声

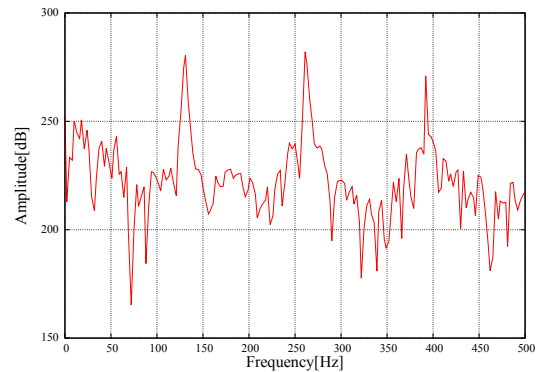


図 5 伴奏音 (ド: 130.8Hz) をフーリエ変換した結果

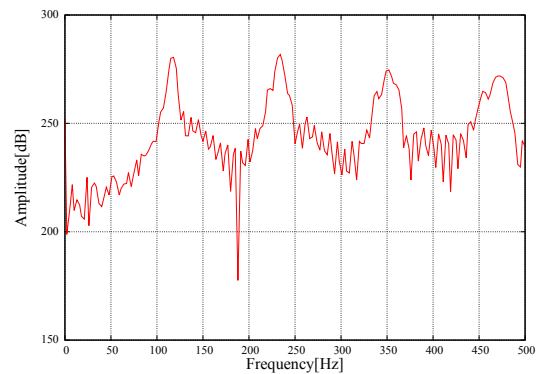


図 6 被験者 A の音声をフーリエ変換した結果

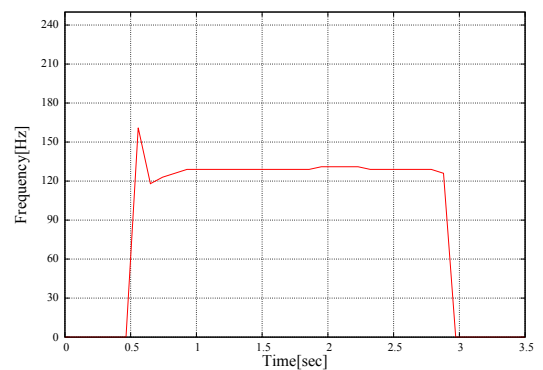


図 7 被験者 A の音程の時間変化

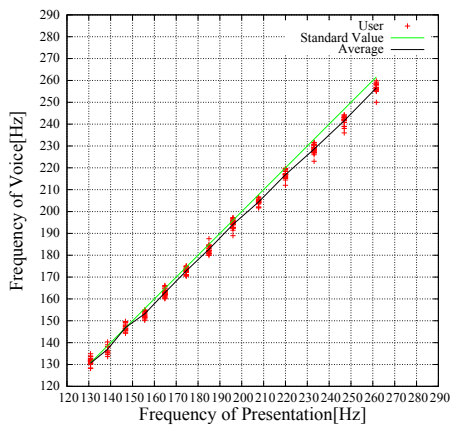


図 8 伴奏音に対する被験者 A の音程

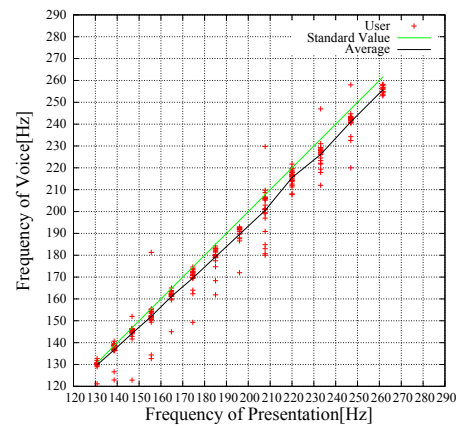


図 11 伴奏音に対する被験者 D の音程

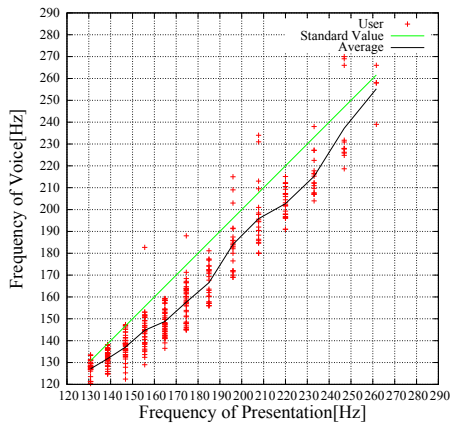


図 9 伴奏音に対する被験者 B の音程

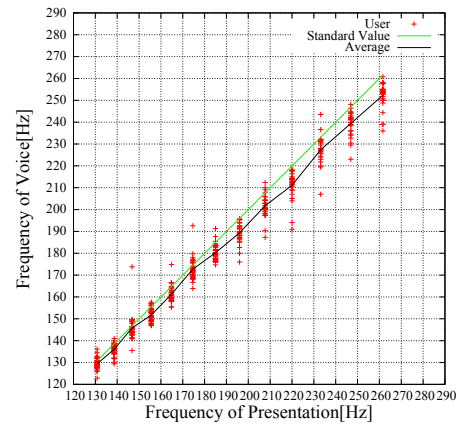


図 12 伴奏音に対する被験者 E の音程

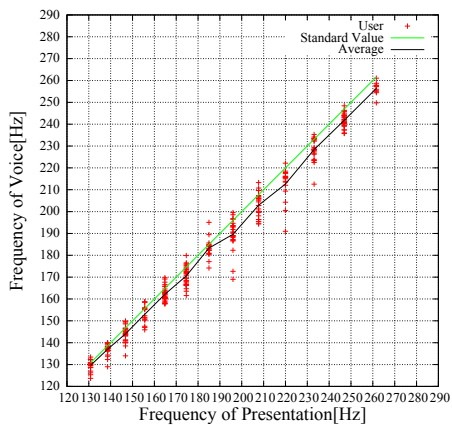


図 10 伴奏音に対する被験者 C の音程

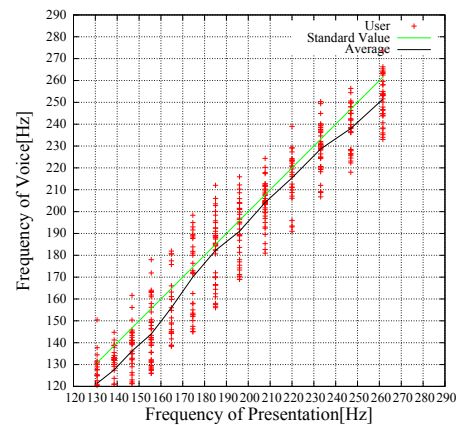


図 13 伴奏音に対する被験者 F の音程

4.3 考察

この結果より、個人によりばらつきの大きさ及び平均値が異なることが確認できた。また、歌唱音の音程の平均が単調増加していることが確認できるため、原曲に比べて不自然な音程の変化をさせることなく音程を操作することが可能であると考えられる。本システムにより音程を操作するためには、ばらつきに比べて、歌唱音の平均値と伴奏音の差が大きい必要があるが、被験者 A はばらつきが小さく平均値が伴奏音よりやや低いため、本システムを用いることにより、音程を正確に操作することが可能であると考えられる。また、被験者 C、D、E においても、ばらつきと、平均値と伴奏音の差がほぼ同程度であり、被験者 B においても、ばらつきの大きさより、平均値と伴奏音の差の方が大きいため、十分音程を操作することが可能であると考えられる。しかし、被験者 F においては、ばらつきの方が、平均値と伴奏音の差より大きいため、今回の単音の伴奏音を提示した際の音程のずれのみについてのキャリブレーションでは、音程を操作することは困難であると考えられる。

5. 対外展示

これまでに、2012 年 9 月 13 日-14 日に慶應義塾大学日吉キャンパスで開催された国際学生対抗バーチャルリアリティコンテスト (IVRC2012) 予選大会において本システムのプロトタイプを実装し、被験者に体験してもらった。この時は前節で述べたようなキャリブレーションは時間的制約で行えなかったため、原曲を聴きながら歌を歌い、その際にキャリブレーションを行い、その結果をもとに、同じ曲を本システムを用いて歌ってもらった。本システムを用いて、30 人程度の人に歌を歌ってもらった。原曲を聴きながら歌った場合と、本システムを用いて歌った場合の音程を原曲の音程との比較を行った。その結果、原曲を聴きながら歌った場合と本システムを用いて歌った場合では、本システムを用いて歌った場合の方が、原曲との音程のずれは少なくなるという傾向が見られた。

6. おわりに

本研究では、「つられる」という現象を利用し、環境音を適切に操作することで歌唱者の音程を意図的に操作し、原曲の音程と同じ音程で歌うことが可能となるシステムを構築することを目指している。今回、単音の伴奏音を提示した際の音程のずれを定量化するため、歌唱者に任意の伴奏音を提示した際の音程のずれを実験的に調べた。その結果、個人によりばらつきの大きさが異なることを確認した。また、単音の伴奏音を提示した際の音程のずれのみについてのキャリブレーションにより、ほとんどの被験者において音程を操作することが可能であるということが確認できた。しかし、一部の被験者においては、伴奏音という要因では歌唱音のずれを決定できないため、音程を操作することは困難であるという結果が得られた。今後は連続音の伴奏音を提示した際の音程のずれを調べることで、伴奏音に対する歌唱音のばらつきが大きい人でも、音程を操作することができるかどうか調べる予定である。

参考文献

- [1] 平井重行, 片寄晴弘, 井口征士. 歌の調子外れに対する治療支援システム. 電子情報通信学会論文誌 D-II, Vol. J84-D-II, No. 9, pp. 1933-1941, 2001.
- [2] 片岡靖景, 伊東一典, 池田操, 中澤達夫, 米沢義道, 今関義弘, 橋本昌己. 歌唱支援システム構築のための歌声の分析と評価. 情報処理学会研究報告 [音楽情報科学], Vol. 98, No. 74, pp. 23-30, 1998.
- [3] 中村隆志. 聴覚妨害時の歌唱実験の解析. 情報処理学会研究報告. 音声言語情報処理, Vol. 97, No. 52, pp. 31-35, 1997.
- [4] 新山王政和. 異なる提示音の間で出現するピッチ知覚の相違に関する実験的研究フラットシンギングとの関係に着目して. 音楽情教育学, Vol. 38, No. 1, pp. 1-9, 2008.
- [5] D.M. Howard and G.F. Welch. Microcomputerbased singing ability assessment and development. *Applied Acoustics*, Vol. 27, No. 2, pp. 89-102, 1989.
- [6] D. Hoppe, M. Sadakata, and P. Desain. Development of real-time visual feedback assistance in singing training: a review. *Journal of Computer Assisted Learning*, Vol. 22, No. 4, pp. 308-316, 2006.
- [7] 川岸基成, 宮島千代美, 北岡教英, 武田一哉. ばね質量系を利用した合唱における歌声の F_0 ダイナミクスのモデル化. 音楽情教育学, Vol. 27, No. 12, pp. 1-6, 2013.
- [8] 加古達也, 大石康智, 亀岡弘和, 永野秀尚, 柏野邦夫, 武田一哉. 合唱における歌声の基本周波数軌跡の分析. 日本音響学会, pp. 1-Q-44(d), 2011.
- [9] 大石康智, 後藤真孝, 伊藤克亘, 武田一哉. 歌声の旋律と動的特徴を特徴付けるための確率的な表現手法に関する検討. 情報処理学会研究報告 [音楽情報科学], Vol. 2007, No. 81, pp. 111-118, 2007.