

## スペクトルの $1/f$ 特性に基づく歌唱合成音声の 音質改善

春木智貴<sup>†</sup> 川端豪<sup>†</sup>

パワースペクトルの  $1/f$  特性は音楽聴取者をより快適にさせることが知られている。本報告では、 $1/f$  スペクトル傾き調整を VOCALOID 歌唱合成音声に適用し、その主観品質への影響を観察する。被験者は 40 秒程度の歌唱合成音声を聴き、42 項目のアンケートに答える。アンケートには、“comfort”, “quality”, “emotion”, “value” などに関係する項目が含まれる。主観評価実験の結果、 $1/f$  スペクトル調整が歌唱合成音声技術に対しても有効であることを確認した。

### Impression of $1/f$ Fluctuated Singing Synthesized Voices

Tomoki Haruki<sup>†</sup> and Takeshi Kawabata<sup>†</sup>

The  $1/f$  fluctuation of the power spectrum makes the music listener more comfortable. This paper describes the improvement of VOCALOID singing voices by using the  $1/f$  spectral compensation. A listener hears the vocaloid songs (about 40 seconds) with three spectrum slope variations, and answers the 42 questionnaire concerning “comfort”, “quality”, “emotion”, and the “value”. Experiments show that the  $1/f$  fluctuation of the power spectrum is also effective for the vocaloid synthesizer technology.

### 1. はじめに

心拍の変動など生体から観測される信号の中には、 $1/f$  の特性を持つものがあることが知られており、人間が快適であると感じる要因の一つにスペクトルの  $1/f$  特性があるとされている[1][2]。  $1/f$  特性とは、パワースペクトルが周波数  $f$  に反比例する特性である。スペクトルの特性について図 1 に示す。横軸が対数周波数、縦軸が対数パワースペクトルの両対数グラフにおける回帰直線の傾き  $k$  を計算しパワースペクトルの傾きを求めることで、スペクトル特性の分析が可能である。

$1/f$  型のスペクトル特性は、他に長期にわたる気温の変動、高速道路上を走る自動車の流量の変動、交響曲の周波数と音の大きさの変動など様々な分野で見ついている[3]。特に、音楽の分野においては、ベートーベン、モーツァルトのピアノソナタなど、人間が聴いて快適に感じる音楽には  $1/f$  スペクトル特性が多く観察される[4]。

このように人間が  $1/f$  特性を示すものに快適と感じる原因として、生体の神経における信号伝達のメカニズムの中に  $1/f$  特性が見られることと関わりがあると考えられている。

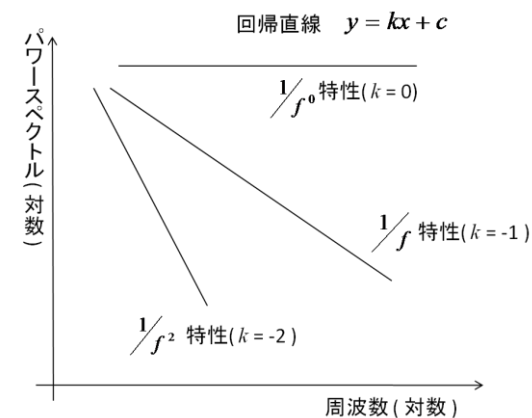


図 1 両対数グラフによるスペクトル傾き

<sup>†</sup> 関西学院大学 理工学研究科  
Kwansei Gakuin University School of Science and Technology

一方で、音声合成技術の発展に伴い、話すだけの合成音声だけではなく、VOCALOID[5][6]などの歌声合成ソフトが開発され、このソフトを利用することで誰でも容易に歌唱合成音声を作成することができるようになっている。VOCALOIDは、歌詞やメロディーを入力するだけで、人の声を元に歌唱音声を合成するソフトであるが、このようにして作成した歌唱合成音声は、人間にとって快適な音声であるとは必ずしも言えない。

本報告では、音楽の分野で効果があるとされている  $1/f$  特性を、歌唱合成音声に対応できるのではないかと考えた。また、主観評価実験によって、歌唱合成音声の快適性の印象について評価を行う。具体的には、人間に心地よい感覚を与えるとされるスペクトルの  $1/f$  特性に着目し、この特性をスペクトルの傾き変換により合成音声に付加することを考える。デジタルフィルタを用いてスペクトルの傾きをコントロールする。このように、歌唱合成音声に  $1/f$  特性を与え、アンケートにより評価を行い、スペクトルの傾きを変化させたことによる印象の変化と、歌唱合成音声に対する  $1/f$  特性の効果について考察を行った。

## 2. デジタルフィルタによるスペクトルの傾き変換

デジタルフィルタを用いて、信号のスペクトル特性を変化することができる。ここでは、デジタルフィルタによるスペクトルの傾きの変換について説明する。1次遅延を含むフローグラフを図2に、2次遅延を含むフローグラフを図3に示す。

$$1 - az^{-1} \quad (1)$$

$$1 - 2az^{-1} + a^2z^{-2} \quad (2)$$

1 次の伝達関数である式 (1) においては、 $a$  の値を変化させることで 6dB/oct までの範囲で、式(2)で表わされる 2 次の伝達関数では 12dB/oct までの範囲で高域を強調することができ、このように簡単なデジタルフィルタを用いてスペクトルの傾きを変化させることができる。

図4に合成音声/aのワースペクトルを、図5に式(2)における2次遅延で得た合成音声/aのワースペクトルを示す。このときの  $a$  の値は1としている。デジタル

フィルタを用いることで、ワースペクトルの傾きほぼ0となっており、ワースペクトルの傾きを変換できていることがわかる。図6に式(1)による1次遅延の周波数特性( $a = 0.8, 1.0, 1.2$ )について示す。この図から、各伝達関数において係数  $a$  を変化させることで周波数特性が変化していることがわかる。以上から、係数  $a$  を適切に設定することによって、歌唱合成音声に対し、任意のワースペクトル傾きを与えることができる。

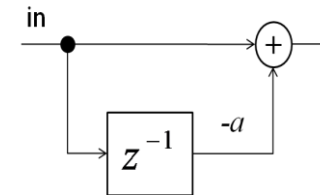


図2 1次遅延を含むフローグラフ

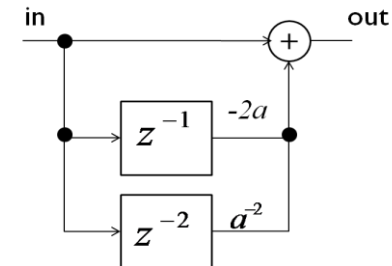


図3 2次遅延を含むフローグラフ

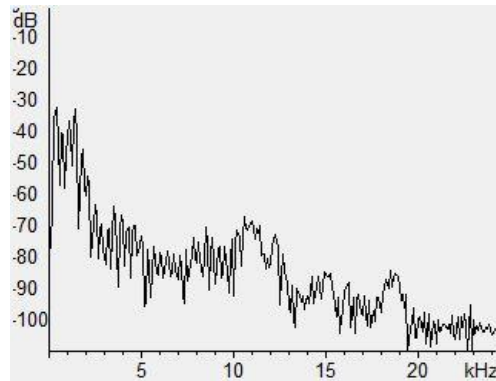


図4 /a/のパワースペクトル

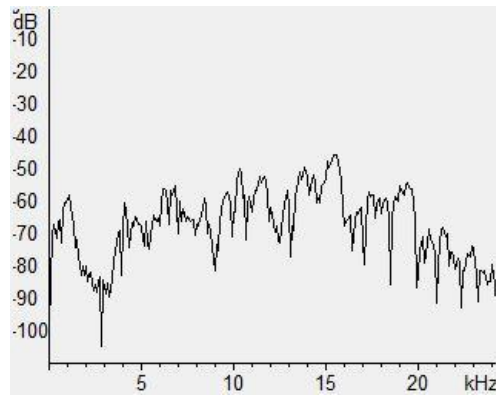


図5 2次遅延で得た/a/のパワースペクトル

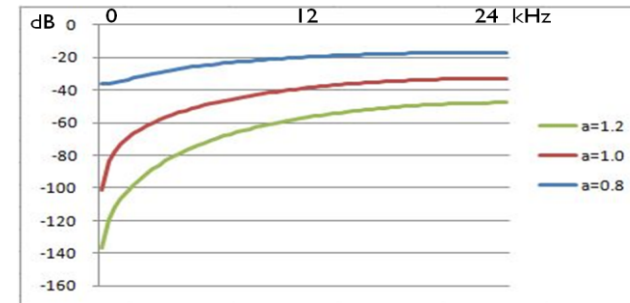


図6 1次遅延の周波数特性( $a=0.8, 1.0, 1.2$ )

### 3. 長時間の曲を用いた評価実験

本報告では、スペクトル傾きの変化がもたらす印象の違いを分析するために主観評価実験を行った。長い時間を聞かせて一括して多数のアンケート項目に答えてもらうため、細かな違いは捉えにくい。ある意味グロスな評価になっている。

#### 3.1 実験条件

実験条件について説明する。実験に用いた歌唱合成音声は、VOCALOID2(初音ミク)を用いて作成した。曲名は Amazing Glace である。被験者は20代の男性8名、女性2名の合計10名とし、後述する42項目のアンケートを用いて、スペクトルの傾きを調整した歌唱合成音声の印象について主観評価実験を行った。

曲の長さは42秒、サンプリング周波数は48kHz、データ形式はリニア16bitである。これに、デジタルフィルタによりスペクトルの傾きを調整した歌唱合成音声2種類を加えた合計3種類を作成した。デジタルフィルタで調整したスペクトルの傾きの値(回帰直線  $k$  の値)は-1.0(1/f特性)と-0.01である。この時、スペクトルの傾き-1.0、すなわち1/f特性をもつ歌唱合成音声は式(1)を用い、傾き-0.01の歌唱合成音声は式(2)を用いて作成した。なお、傾き調整処理を行っていない歌唱合成音声のスペクトルの傾きは-2.0となっている。実験環境は、暗騒音33.7dBAの室内、被験者にはヘッドホン(SENNHEISER HD280pro)を用いて刺激音を両耳に呈示した。アンケートの項目(42項目)は、Osgood.C.Eの提案したSD法による形容詞対から抜粋したものをを用いた。また、アンケートのスコアについては(かなり、やや、どちらかといえば、どちらでもない、どちらかといえば、やや、かなり)とし、それぞれ(-3, -2, -1, 0, +1, +2, +3)の全7段階に設定した。

評価実験の際に、音量による主観評価への影響を避けるため、実験を始める前にテストデータを再生しながら被験者が各自で音量の設定を行うようインストラクションにて指示を行った。音量の設定後、アンケート項目全てに目を通してもらい、曲の再生中は裏返しにするよう指示した。曲の印象を忘れたり混ざったりするのを防ぐため、アンケート全体の解答時間は約2分に設定し、1曲聴き終える毎に直観的に解答するよう指示した。アンケート用紙はあらかじめ印刷したものを提示曲数分である3枚用意し、筆記にて解答させた。

### 3.2 実験結果・考察

被験者10名のアンケートの解答結果について、全42項目それぞれについて分散分析を行い、有意水準1%と5%において有意差のある項目を調べた。表1に有意水準1%で有意とされた11項目、表2に有意水準5%で有意とされた6項目を示す。これらは観測された分散比の大きい項目から順に並べている。1列目に観測された分散比、2列目にアンケート項目の形容詞対、3列目にそれぞれの項目の形容詞が属するおおまかなカテゴリとして Comfort (快適性), Value (価値), Quality (品質), Emotion (感情) のいずれかを筆者の主観で与えた。図7に表1、表2で挙げられた項目について被験者10名のアンケート解答結果のスコアの平均を示す。横軸はアンケートの項目、縦軸はアンケートのスコアとなっており、スペクトルの傾き3種類がそれぞれ記号で示されている。なお、項目(眠い-眠くない), (柔らかい-堅い), (おだやかな-興奮する)についてはスコアの値が低いほど、評価が良く、それ以外の全ての項目ではスコアが高いほうが良い評価を得たものとしている。

水準1%で有意とされた項目に(耳に快い-不協和な), (リラックスできる-緊張する), (興奮する-おだやかな)といった快適性と設定した Comfort に分類される項目が上位に多く含まれている。図7より、これのアンケート項目は全て共通してスペクトルの傾き-1.0のときに最も評価が高く、傾き-2.0, -0.01の順に評価が悪くなっている。このことから、従来から考えられていることだが、スペクトルが1/f特性を示すものに人間は快適と感じる傾向が、歌唱合成音声にも表れていることがわかった。また、価値観と設定した Value, 品質と設定した Quality といった項目においても Comfort と同様にスペクトルの傾きが-1.0のときに最も高い評価が得られ、傾き-0.01のときに最も評価が低くなっている。このことから、スペクトルの1/f特性は Comfort だけではなく、Value や Quality にも有効であることがわかった。

次に、水準5%で有意とされた項目には(ありがたい-恩知らずの), (眠くない-眠い), (明るい-真っ暗な)のような感情と設定した Emotion に分類される項目が表れている。1/f特性の効果として、従来から有効とされていた Comfort 以外にも、今回の実験から Emotion にも効果あるということがわかった。

以上の検討から、スペクトルの傾きを調整することで、傾きが-0.01のときに最も評

価が低く、傾きが-1.0のときに最も高い評価を得るとわかった。しかしながら、図7を見てもわかるとおり、アンケートの解答結果の差が1前後となっており、曲の時間が長いことやアンケート項目が多いことから、これ以上傾きのパターンを増やしても詳細な結果が得られ難いと考えられる。そこで、スペクトルの傾きの違いがもたらす効果についてより詳細な分析を行うため、曲の時間とアンケート項目を少なくし、対比較による評価実験を行った。次節で述べる。

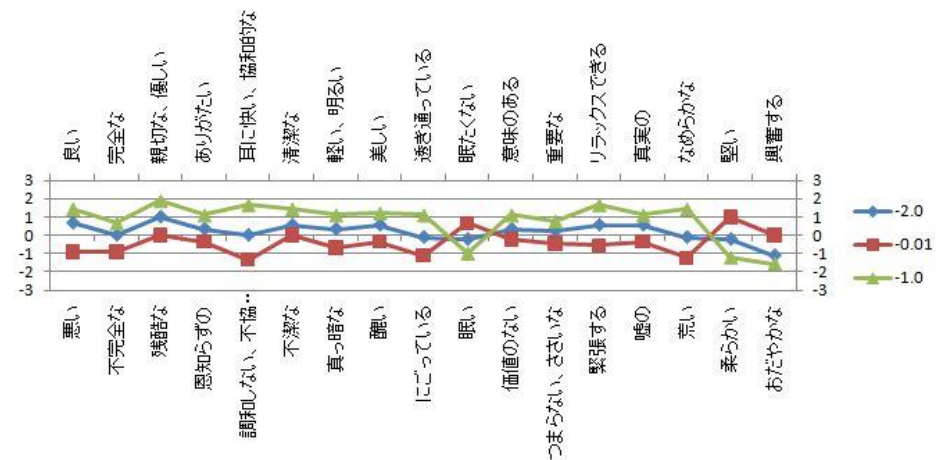


図7 集計結果平均

表1 1%で有意

分散比	アンケート項目		属性
23.6	耳に快い, 協和的な	調和しない, 不協和な	Comfort
12.7	良い	悪い	Value
9.9	リラックスできる	緊張する	Comfort
8.7	親切な, 優しい	残酷な	Comfort
7.7	なめらかな	荒い	Quality
7.2	興奮する	おだやかな	Comfort
7.0	堅い	柔らかい	Quality
6.8	意味のある	価値のない	Value
6.7	透き通っている	にごっている	Quality
6.4	真実の	嘘の	Value
6.2	美しい	醜い	Quality

表2 5%で有意

分散比	アンケート項目		属性
4.8	ありがたい	恩知らずの	Emotion
4.6	眠たくない	眠い	Emotion
4.5	重要な	つまらない	Value
4.0	清潔な	不潔な	Value
3.8	完全な	不完全な	Value
3.7	明るい	真っ暗な	Emotion

#### 4. 短時間の曲を用いた評価実験

スペクトルの傾きの違いがもたらす効果について、より詳細な評価を行うため、短い時間の曲を聞かせ前節で有効性の高かった5項目について対比較実験を行った。

##### 4.1 実験条件

対比較実験に使用した歌唱合成音声は、3節の評価実験の際に用いた Amazing Grace である。ただし、曲の長さは冒頭の約5秒とした。デジタルフィルタによりスペクトルの傾きを-0.5と-1.5に調整した新たな歌唱合成音声2種類を作成し、3節で用いた傾き-1.0、-0.01、-2.0と合わせて全5種類の歌唱合成音声を使用した。サンプリング周波数は48kHz、データ形式はリニア16bitである。現時点での被験者は20代の男性1名である。アンケート項目には、3節の実験により得られた1%で有意とされた表1の項目から、上位5つの項目左側の形容詞（耳に快い、良い、リラックスできる、親切な、なめらかな）を用いた。実験環境は、暗騒音33.7dBAの室内で、使用したヘッドホンは SENNHEISER HD280pro である。

対比較実験にはサーストン法を基にし、組み合わせ数はスペクトルの傾き5種類（-0.01、-0.5、-1.0、-1.5、-2.0）の全10パターンとなっている。呈示された2曲を1回ずつ聴いてもらい、両者を比較した際に5つのアンケート項目から当てはまる項目をどちらか1つに必ず選択させ、選択された項目数を用いて順位付けを行った。5項目のうち多くの項目が選ばれた方を評価の高いものとする。アンケートには呈示された2曲を聴いた後即座に解答させた。アンケート用紙はあらかじめ全組み合わせである10枚用意しておき、筆記で解答させた。

##### 4.2 実験結果・考察

全5種類の呈示曲についてサーストンの対比較法を基に順位付けを行った。結果を図8に示す。横軸がアンケート項目の選択率から求めた標準正規分布の逆関数の列平均、横軸上にプロットされた点上の数値がスペクトルの傾きである。左から右に行くにつれて評価が低いものとなっている。この図から、スペクトルの傾き-1.5のときの評価は-1.0と-2.0の間、スペクトルの傾き-0.5のときの評価は-1.0と-0.01の間になり、良いと評価されたスペクトルの傾きの順は、-1.0、-1.5、-2.0、-0.5、-0.01となっている。この結果から、スペクトルの傾きが-1.0のときに最も高い評価を得ることがわかった。スペクトルの傾きを-2.0から-0.01の間で変えると、一度評価は上がり-1.0でピークを持ち、その後評価は下がることがわかる。

これらの実験により、従来より言われてきたように、スペクトルの1/f特性が聴取者の快適性に貢献するという知見が歌唱合成音声に対しても成立していることを確認できた。

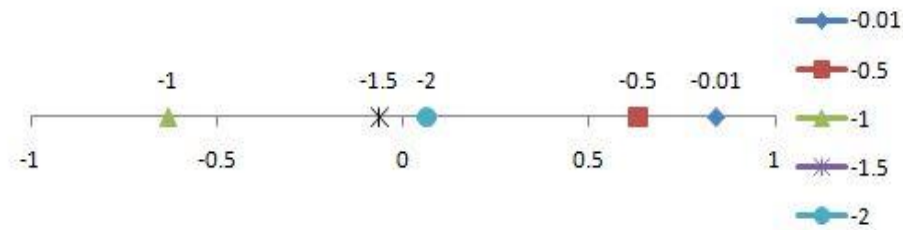


図 8 対比較実験による順位付け

## 5. 結論

本研究では、スペクトルの傾きを変化させた歌唱合成音声の印象と、スペクトルの  $1/f$  特性の効果について実験、考察を行った。デジタルフィルタを用いてスペクトルの傾きをコントロールし、アンケートを用いて主観評価実験を行った。スペクトルの傾きが  $-1.0$ 、すなわち  $1/f$  特性を与えた歌唱合成音声に対して Comfort に分類したアンケート項目の評価が高く、Value や Quality といった面でも  $1/f$  特性は効果をもたらすことがわかった。傾きの変化がおよぼす効果についての詳細な分析を行うため対比較実験を行った結果、好印象な曲のスペクトルの傾きは  $-1.0$ 、 $-1.5$ 、 $-2.0$ 、 $-0.5$ 、 $-0.01$  の順で表現でき、傾き  $-1.0$  のときに最も良い評価を得ることがわかった。従来から述べられていた  $1/f$  特性の快適性に関する効果は歌唱合成音声にも有効であることが確認できた。今後は対比較実験における被験者数を増やし、より詳細な分析を行っていく。

## 参考文献

- 1) 武者利光: 生体情報と  $1/f$  ゆらぎ, 応用物理, vol.54, pp429-435 (1985)
- 2) 武者利光:  $1/f$  ゆらぎと快適性, 日本音響学会誌 vol.50, pp485-488 (1994)
- 3) 武者利光:  $1/f$  ゆらぎ, 応用物理, vol.46, pp1144-1155 (1977)
- 4) 佐治晴夫:  $1/f$  ゆらぎ-ヒトと宇宙をつなぐもの, 日本ファジィ学会誌, vol.6, pp620-632 (1994)
- 5) 剣持秀紀, 大下隼: 歌声合成システム VOCALOID, 情報処理学会研究報告, Vol.2007-MUS-72, pp.25-28 (2007)
- 6) 剣持秀紀, 大下隼: 歌声合成システム VOCALOID-現状と課題, 情報処理学会研究報告, Vol.208-MUS-74, pp.51-56 (2008)