

音色を対象とした主観評価モデルの動的生成機構を有する音響合成システム

陳 昀劭[†] 清木 康[†]

[†] 慶應義塾大学環境情報学部 〒252-0882 神奈川県藤沢市遠藤 5322

E-mail: [†] {t16577yc, kiyoki}@sfc.keio.ac.jp

あらまし 音声は、人間の感性を介して、多様な情景を思い浮かべる。言語情報を含めない単純な音を与える印象は、音色によって左右される。しかし、同じ音色にしても受け取る印象は聴者によって異なるため、感性的特徴の解析に際して、聴者の感性に応じた調整が必要となる。本稿では、聴者による印象評価に基づいた、音色に対する感性を反映させる主観評価モデルの動的生成機構、そして同機構を搭載した音響合成システムの実現方法を示す。主観評価モデルの適用により、聴者それぞれの感性に合致した音色情報を提供することで、本システムを通して繊細な個人差を具現化することが実現される。

キーワード 音声合成, 感性情報処理, 感性工学, データモデリング, 可視化

1. はじめに

人類は、古き時代から音に興味を抱いていた。音響学の起源をたどると、古代ギリシャのピタゴラスにまで遡ることができる。また、最古とされている楽器の出土から、演奏が長らく行われてきたことがわかる。ピタゴラスから 2000 年以上が経過した今でも、音の謎は追究され続けている。

日常生活において、様々な音が我々を取り巻いている。しかし、カクテルパーティー効果が観察されるように、同時に存在するあらゆる音の正体を理解することはほぼ不可能である。音は、身近で遥かなものであった。音響機器の普及に伴い、人間は既存の楽器にとどまらず、時空の制限を超えて心地よい音を追求するようになった。また、コンピュータの出現により、誰でも手軽に音の研究に着手することが可能になった。

音色に関する研究は最近、より多角的になされている。従来の音響工学、信号処理に加え、音楽情報処理、デジタルアート、音楽社会学など、比較的に新たな分野においても盛んである。音色が与える印象がどのような要素に依存しているかは、これまで少なからず研究されてきた。多数の先行研究において実験を通して得られた因子が多様である点から、不変な因子群を不特定多数の聴者に適用することに比べ、聴者各々のモデルを構築した上で因子を特定することの精度がより高いと予想される。

本研究は、感性情報マネジメントの視点から音色を探究する。音色の印象にまつわる因子について検証しつつ、データモデリング手法の提案、そして感性の可視化・可聴化を試みる。

2. システム構成

図 1 に、本研究で開発したシステムの構成図を示す。各機能の動作原理については、3.3 節～第 4 章に示されている。

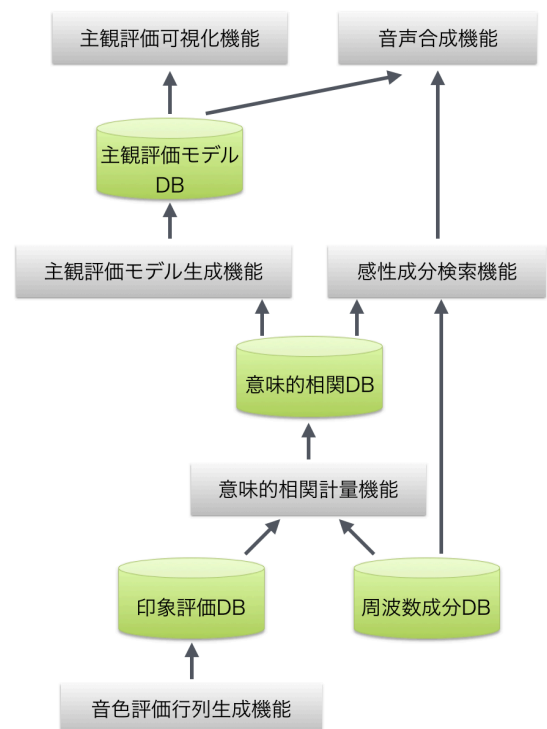


図 1 システム構成

3. 音色に関する印象評価の計量系

本システムは、感性における個人差を音響合成に反映するために、ユーザによる印象評価を取得する。以下、音色評価に関する先行研究および、それらを踏まえた本研究で採用する評価方式を示す。

3.1 先行研究

音色の印象評価に用いられる手法として、言葉とそれぞれの度合いにて記述することが普遍的である。しかし、印象を表す形容詞、時には名詞は数え切れないほど存在するため、印象評価を必要とする実験では、なるべく少数の言葉であらゆる印象を網羅できるように措置することが必須となる。すなわち、音色評価に際して印象語の選別は大きな課題である。このように、評価尺度を体系化しなければならないという需要に応じて、これまで多くの研究が行われてきた。

評価に用いられた言葉にこそ相違があるものの、因子分析の結果を比較すると、音色の印象を表す印象語は3~4因子に集約できることが明らかになった。[1] その代表例を表1に示す。

表1 音色評価に関する先行研究

発表者	因子
北村, 二井, 栗山, 増田 [2]	美的因子、迫力因子、金属性因子、(柔らかさ因子)
曽根, 城戸, 二村 [3]	美的・叙情的因子、量的・空間的因子、明るさ因子、柔らかさ因子
von Bismarck [4]	sharpness 因子、compactness 因子、fullness 因子、colorfulness 因子

北村らは、3 因子および時おり観察された柔らかさ因子を抽出できたことを発表した。曽根らの実験で得られた因子は、呼称が異なっているものもあるが、各因子と相関の強い印象語を見ると共通するものが多く含まれているため、北村らが示した因子とほぼ同様と言える。録音を用いた上述2件の研究に対して、合成音を用いた手法の代表例として von Bismarck の実験があげられる。感性における個人差を音響合成によって表すという本研究の目的を念頭に置き、本システムは von Bismarck の評価手法を採用する。

3.2 von Bismarck による音色評価実験

von Bismarck は、30 語の形容詞およびその対義語で構成された評価尺度を用いた。実験参加者は上記の尺度で35種類の定常音に対して印象評価を行った。使用

された定常音は、周波数スペクトルが 200Hz~8kHz の範囲に収まるように合成されたものである。また、調波複合音およびノイズの両方が含まれていた。

音楽経験者と未経験者との人数が均等となるように評価を集計し、この印象評価に対して因子分析を行った結果、評価全体の分散が4因子によって説明されることが判明した。表1に示されているように、4因子は負荷の高い評価尺度にて命名された。

3.3 印象評価の取得

本システムはまず最初に、von Bismarck の実験に用いられた音声を再現し、サンプル音声としてランダム順に提示する。システムを使用するユーザは個々の音色について評価を行う。ただし、von Bismarck の実験で刺激音の一部として提示されたドイツ語の母音に関しては、言語的側面の影響による偏りが予想されるため、本研究はこれら母音の音声をあらかじめ排除する。印象評価を集計することで、ユーザの個人性をシステムに反映する。

音色に対するユーザの印象評価は、音色評価行列生成機能によって行列データに変換され、印象評価データベースに格納される。

4. システムの実現方式

ユーザの感性に基づいたパーソナライズを行った後、意味的相関計量や主観評価モデル生成などのプロセスを踏むことで音響合成に必要なデータを抽出することが可能になる。本章では、その一連の処理において各機能がいかに動作するかについて述べる。

4.1 感性因子と周波数成分との意味的相関

本システムにおける意味的相関計量機能は、意味的連想検索方式[5][6]を利用して実装された。

意味的連想検索方式とは、検索対象に複数の形式にて記録されたデータが含まれている場合に、検索が実行される状況を考慮した上で検索キーワードと検索対象との意味的な関係を計算し、マルチメディアに対応可能な情報検索方式である。なお、「検索が実行される状況を考慮する」点は、同方式では文脈理解と呼ばれる。

音を聴くと情景が思い浮かぶ、また多様な音色は多数の波が合成された結果、この2点は、音声に感性的側面および物理的側面があることを示唆している。そして、感性と物理量は前述の文脈であると解釈することもできる。こうした特質に注目し、本研究では意味的連想検索方式の拡張用法として意味的相関を計量し、その結果を用いて感性の可視化・可聴化を実現する。

音色評価実験における評価尺度としての印象語群は、

音色の感性に関連する要素を表しており、また因子分析にて集約することができる。この 2 点から、本稿では印象語群で表しうる、感性に付随した要素を感性因子と称する。意味的連想検索方式によって、各サンプル音声は感性と物理量との 2 文脈に射影される。2 文脈間の意味的相関を計量するため、本システムでは感性因子および周波数成分をデータ分析に用いる。感性因子は印象評価データベースに保存されている行列データから計算可能である。一方、各サンプル音声の周波数成分は周波数成分データベースに格納されている。

意味的相関計量機能は上記 2 つのデータベースからデータを読み出し、意味的相関を計量する。印象評価データベースに保存されているデータは、von Bismarck の評価方式を採用しているため、7 段階評価にて記述されている。ユーザによって評価値の分布にかなりなばらつきが予想されるため、データ前処理として正規化を行う。ユーザ別の評価結果の分布が異なるが、評価尺度が同一である故最大値と最小値は定まっている。したがって、Min-Max Normalization が適している。印象評価の原数値を V 、変換後の数値を V' 、評価尺度における最大値を V_{max} 、最小値を V_{min} 、中央値を V_{med} とすると、正規化は以下の式を満たす。

$$V' = \frac{V - V_{med}}{\frac{1}{2}(V_{max} - V_{min})}$$

また、7 段階評価の場合には、 $V_{max} = 7$, $V_{min} = 1$ という条件も満たされる。von Bismarck の評価方式は対義語からなる両極尺度を使用しているため、 $[0, 1]$ ではなく、変換後の数値の範囲が $[-1, 1]$ となるように正規化を行う。正規化で得られたデータは、実際に意味的相関の計量に用いられる感性因子行列となる。計量に必要なもう一つのデータセットは、各サンプル音声の周波数成分を記録した周波数成分行列である。

[0.666667, 0.666667, 0.333333, 0.666667, -0.333333,
0.000000, 1.000000, 0.666667, -0.666667, 0.666667,
-0.333333, 0.666667, -1.000000, 1.000000, 0.333333,
0.666667, 0.333333, -1.000000, 0.000000, 0.333333,
-0.333333, 0.333333, -0.333333, -1.000000, -0.666667,
0.000000, -1.000000, 0.333333, 0.666667, 0.333333]

図 2 感性因子行列

Sample	200	201	202	203	204
B1	1	0	0	0	0
B2	1	0	0	0	0
B3	1	0	0	0	0
B4	1	1	1	1	1
B5	1	0	0	0	0
B6	1	0	0	0	0
B7	1	1	1	1	1
B8	1	0	0	0	0
B9	1	0	0	0	0
B10	1	0	0	0	0
B11	1	0	0	0	0
B12	1	0	0	0	0
B13	1	0	0	0	0
B14	1	0	0	0	0
B15	1	0.9931	0.9863	0.9795	0.9727

図 3 周波数成分行列（一部）

表 2 印象語群

hard-soft
sharp-dull
angular-rounded
obtrusive-reserved
unpleasant-pleasant
tense-relaxed
loud-soft
violet-gentle
bright-dark
strong-weak
high-low
rough-smooth
restless-calm
complex-simple
coarse-fine
dirty-clean
compact-scattered
boring-interesting
narrow-broad
closed-open
dead-lively
tight-wide
ringing-dampened
pure-mixed
thin-thick
empty-full
colorless-colorful
dim-brilliant
solid-hollow
heavy-light

図 2 に感性因子行列、図 3 に周波数成分行列の一例が示されている。感性因子行列における個々の要素に対応する印象語は、表 2 の通りである。右側の印象が強いほど数値が大きくなる。

感性因子行列と周波数成分行列、この 2 点のデータを分析することで意味的相関を算出する。音色評価の取得は、各感性因子が独立して作用することを前提に実施された。この点から、意味的相関の計量には重回帰分析が相応しいと考えられる。各周波数成分の振幅からなる説明変数および、感性因子の強度である目的変数によって重回帰モデルが構成される。サンプル音声を構成する波形の最低周波数を b Hz、上限周波数を u Hz、周波数 i Hz の振幅を A_i 、偏重回帰係数を C_{Ei} 、各感

性因子の強度を K とすると、重回帰モデルは以下の式で表すことができる。

$$K = C_S + \sum_{i=b}^u C_{Ei} A_i$$

(C_S : 定数項)

重回帰分析にて得られた偏回帰係数は、周波数成分と感性因子との意味的相関を表現した数値である。意味的相関計量機能は、各感性因子の強度を偏回帰係数および定数項にて改めて記述し、意味的相関の行列データを出力する。このデータは意味的相関データベースに格納される。

意味的相関データベースにアクセスするもう一つの機能は、感性成分検索機能である。感性成分検索機能は感性因子を用いた検索を支援する。意味的相関データベースへの問い合わせでもって、入力された各感性因子の強度を周波数成分個々の偏回帰係数および定数項に分解する。このように、感性成分検索機能は各感性因子が含まれる比率に基づいて周波数スペクトルを算出し、音色情報を返す。この情報は、音声合成機能に用いられる。

4.2 音色に対する主観評価モデルの生成機構

主観評価モデルとは、ユーザの印象評価データを分析することで得られた、同ユーザが重視する感性因子の集合をモデリングしたものである。

本システムは von Bismarck の評価方式を使用しているため、同様に因子分析によって重要な感性因子を洗い出す。von Bismarck は、実験で集計したデータの分散の 91% を 4 因子で説明できることを発表した。ところが、必ずしもあらゆる人々の印象評価データに 4 因子が最適だとは限らない。ユーザの個人性に着眼した本システムでは、因子数は可変である。感性因子の強度を観測変数にし、共通因子およびそれぞれの因子負荷を導出する。

上記の処理は主観評価モデル生成機能によって実行される。閾値の設定によって、共通因子がどの感性因子に強く関係しているかを確認することができる。図 4 は観測変数が受けている各共通因子の影響を示したものである。同データの結果から、共通因子各々の負荷が高い観測変数を図 5 に示す。(因子数 = 4, 閾値 = 0.6)

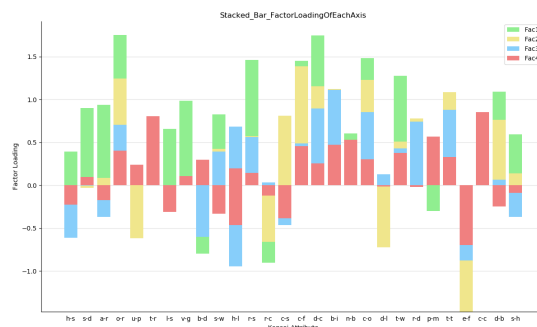


図 4 各観測変数に対する共通因子の影響

Attributes: Corresponding To Each Factor
(Loading >= 0.6)

Factor 1 : ['hard-soft', 'sharp-dull',
'angular-rounded', 'loud-soft',
'violet-gentle', 'rough-smooth',
'complex-simple', 'tight-wide']

Factor 2 : ['coarse-fine', 'compact-scattered',
'dim-brilliant', 'heavy-light']

Factor 3 : ['strong-weak', 'dirty-clean',
'boring-interesting',
'ringing-dampened']

Factor 4 : ['tense-relaxed', 'colorless-colorful']

図 5 各共通因子を反映した観測変数

ユーザ個人の主観評価モデルは、因子負荷の平均値によって構築される。ただし、因子負荷は[-1, 1]の範囲に分布するため、ポジティブに作用する因子とネガティブに作用する因子を 2 グループに分ち、計算を行う。主観評価モデル生成機能によって算出されたデータは、主観評価モデルデータベースに格納される。同データベースにアクセスし、主観評価可視化機能はグラフ化した主観評価モデルを出力する。図 6 に主観評価モデルの一例を示す。緑色が正数の因子負荷の平均値であり、一方で赤色は負の因子負荷の平均値を示している。また、これは図 4、図 5 と同ユーザのデータを分析した結果である。

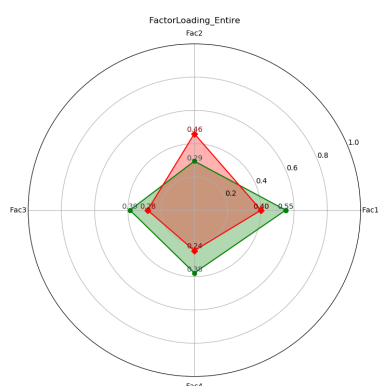


図 6 主観評価モデル

主観評価モデルにより、印象評価を行ったユーザが目にした感性因子群、さらにそれらが潜在的な共通因子にいかに関与されているかについて可視化することが可能となる。

4.3 主観評価モデルを用いた音声合成機能

音声合成機能は、指定された感性成分を表現する音声データを出力する。

感性因子と周波数成分との意味的相関に基づき、感性成分検索機能は感性因子強度を物理量の文脈に射影した結果、つまり対応する周波数成分を抽出する。この情報が音声合成機能に渡され、実際の音声データとして出力される。

音声合成に際して、主観評価モデルを参照し、因子負荷が低い感性因子の投影によって発生した周波数成分を除却することも設定できるように音声合成機能が実装されている。こうしたフィルタリングを通して、本システムでは、ユーザにとって重要な感性因子のみを音色の形式で再構築することが実現可能である。

5. 実験・考察

本システムを使用し、実際に音色評価データを分析した結果を基にして考察を行う。前述したように、本研究は個人差に着目している。印象評価の一貫性を保つために、本章で示した実験結果は、同一ユーザの印象評価によるものである。(図 6 を含める)

5.1 von Bismarck の 4 因子に関する検証

本節では、共通因子数を変更し、主観評価モデル生成機能が出力した因子負荷の平均値を比較することで、von Bismarck が提案した 4 因子の妥当性について検証する。

4 因子の結果はすでに図 6 に示した通りである。図

7 に 5 因子、図 8 に 6 因子にて因子分析を行った際の主観評価モデルを示す。

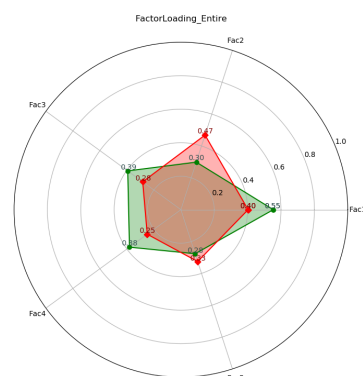


図 7 主観評価モデル (5 因子)



図 8 主観評価モデル (6 因子)

5 因子での分析結果から明らかなように、第 5 因子 (Fac5) によるポジティブな影響とネガティブな作用は近似した値を有している。言い換えれば、第 5 因子の挙動は比較的に不明瞭であり、共通因子として解釈することがほかの 4 因子より困難である。また共通因子数を 6 に設定すると、第 6 因子 (Fac6) の負荷があらゆる感性因子軸において 0 となっており、すなわち、第 6 因子が存在しないことが判明した。この結果は、von Bismarck の論述を支持し、音色評価の分析に 4 つの共通因子を用いた手法の一般性を示唆している。

5.2 意味的相関計量の 2 モデル

意味的相関計量機能の開発に際して、本研究は 2 通りのモデリング手法を考案した。以下、意味的相関および音響合成の視点から双方の性能について比較する。

モデル A

重回帰分析において、各感性因子の強度を説明変数にし、周波数成分の振幅を目的変数にする。感性成分検索機能は、指定された各感性因子の強度と意味的相関にて周波数成分個々の振幅を計算し、周波数スペクトルを出力する。この場合、意味的相関を表した数値は各感性因子の偏回帰係数である。

モデル B

重回帰分析において、各周波数成分の振幅を説明変数にし、感性因子の強度を目的変数にする。感性成分検索機能は、指定された各感性因子の強度と意味的相関にて周波数成分個々の振幅を計算し、周波数スペクトルを出力する。この場合、意味的相関を表した数値は各周波数成分の偏回帰係数である。またこのモデルは、すでに 4.1 節に示したものである。

感性成分検索機能の動作が同然に見えるが、意味的相関が異なる数値によって表現されている。これは、文脈間にて射影を行う順序の相違による結果である。モデル A において、周波数成分が感性因子群によって構成されるため、入力された感性因子の強度から直接、周波数スペクトルを算出することができる。一方、モデル B では周波数成分の集合が一つの感性因子を構成するため、検索の際にはまず指定された感性因子強度に応じて周波数成分が同じ比率にて増幅・減衰され、各感性因子が提供する周波数成分を合成した結果が出力される。このように、モデル A は意味的相関を計量する際に感性の文脈に射影を行い、反対にモデル B は感性成分検索機能の作動中に物理量の文脈への射影を行う。

印象評価の原数値を用いた場合には、双方ともノイズ寄りの音声を出力した。4.1 節に示した正規化を導入した結果、モデル B は明晰な音色を合成し、モデル A では問題点は依然として改善されなかった。この点から、本研究の目的に相応しいデータモデリング手法はモデル B だと推測できる。

6. 結論

本稿を通して、主観評価モデルの動的生成機構およびそれを用いた音響合成手法を示した。

von Bismarck の実験方法を再帰的に利用することで、音色に対する感性における個人性を具現化することが可能となった。また、一歩進んだ試みとしては、感性因子と周波数との関連性を分析し、本システムを使用するユーザの印象評価に応じてダイナミックにデータモデリングを行う点である。

意味的相関を計量した上でデータを抽出する方式

は、従来のデータベースシステムに多用されるパターンマッチングに比べ、より緻密に情報を管理・提供できる。さらに、メタデータのパーソナライズを通して、ユーザは各自の感性でもって問い合わせることが可能である。本方式は、本研究で対象としている音声に限らず、感性を引き起こしうる様々なメディアへの応用が期待される。

参 考 文 献

- [1] 岩宮眞一郎, 日本音響学会, "音色の感性学: 音色・音質の評価と創造" コロナ社, 2010
- [2] 北村音一, 二井真一郎, 栗山譲二, 増田昇, "昭和 50 年代の青年に関する音色因子の抽出" 日本音響学会聴覚研究会資料 H-51-11, 1978
- [3] 曽根敏夫, 城戸健一, 二村忠元, "音の評価に使われることばの分析" 日本音響学会誌 Vol.18-6 p.320-326, 1962
- [4] G. von Bismarck, "Timbre of Steady Sounds: A Factorial Investigation of its Verbal Attributes" Acustica Vol.30, p.146-159, 1974
- [5] 清木康, 金子昌史, 北川高嗣, "意味の数学モデルによる画像データベース探索方式とその学習機構" 電子情報通信学会論文誌 Vol.j79-D-2 No.4, p.509-519, 1996.
- [6] 吉田尚史, 清木康, 北川高嗣, "意味的連想検索機能を持つメディア情報検索システムの実現方式" 情報処理学会論文誌 Vol.39-4, p.911-922, 1998
- [7] 難波精一郎, "音色の定義を巡って" 日本音響学会誌 Vol.49-11, p.823-831, 1993
- [8] Matplotlib でレーダーチャートを描く (16 行)
<https://qiita.com/1007/items/80406e098a4212571b2e>
(2019.12.30 閲覧)