

# 画像と音の調和に関する心理的な分析とそのモデル化

## -周波数による二音和音の協和・不協和の数理モデル化-

柴田 滝也<sup>†</sup>

† 東京電機大学情報環境学部 〒270-1382 千葉県印西市武西学園台 2-1200

E-mail: shibata@sie.dendai.ac.jp

**あらまし** 視覚情報と聴覚情報の調和・不調和を心理的な観点より、分析・モデル化を行うとともに、映像と音楽のマッチングを支援し、あるいは、心理的にも不快に感じるコンテンツを排除するシステムを構築することを目的とする。本稿では、Plomp & Levelt による二音和音の協和・不協和と周波数との関係を画像コントラストの仕組みを参考にしてモデル化を行った。将来的には、効果音や音楽の周波数特性を分析し、そのモデルを用いて協和・不協和を分類し、心理レベルでの画像と効果音や音楽の調和・不調和の仕組みの分析およびモデルを構築することを目指す。

**キーワード** コントラスト、協和・不協和音、マルチメディア、調和、モデル化

## Analysis and modeling of harmony between image and sound

### -Frequency model of consonance and dissonance tone-

Tatsuya SHIBATA<sup>†</sup>

† School of Information Environment, Tokyo Denki University 2-1200 MuzaiGakuendai, Inzai, Chiba,  
270-1382 Japan

E-mail: shibata@sie.dendai.ac.jp

**Abstract:** The purpose of this paper is to construct the environment that provides user-oriented information by psychologically analyzing and modeling harmony between vision and sound information. We have modeled the harmony between image and sound on the psychological level by analyzing not only two piano tones but also the both. The paper shows that a consonance and dissonance model of two piano tones based on the contrast model related to the image contrast. The distribution of the model is similar to the one by Plomp and Levelt. It seems that the contrast is related to not only vision information process but also sound one.

**Keyword:** Contrast, Consonance & dissonance Tone, Multimedia, Harmony, Modeling

### 1. はじめに

デジタル放送などによるマルチメディア・コンテンツの需要が増加し始めている。コンテンツ作成時には映像と音声・音楽のマッチングなどの編集に多くの労力と時間を費やしている。そこで視覚情報と聴覚情報の調和の仕組みがわかればインターネット上にある情報から選択することが可能になり、そのような編集作業は軽減される。本研究では映像と音・音楽のマッチングを支援し、また、心理的に不快に感じるコンテンツを排除するシステムを構築するためのシステムを研究することを目的とする。そのために、心理レベルでの画像と音の・不

調和の仕組みの分析を行い、画像と音の相互作用を考慮したモデルを構築する。そのモデルから音に画像を、あるいは、画像に音を付加することによって、相乗的に心理的效果をあげる手法を構築する。

調和に関する研究としては、画像の二色配色の調和についての分析がある。ムーン・スペンサーは二色配色の色彩調和の理論[1]を提案し、色相の「同等」、「類似」の二色配色では調和度が高くなる一方、少し異種の色相では調和度が低くなることを示した(図1)。この現象はすべての色相に相対的に起こるとしている。

一方、同じような現象が二音和(純)音にもあ

る。Plomp & Levelt[2]によると、異なる周波数からなる二音和音において、その周波数との差によって、協和・不協和が生じると述べている(図2)。

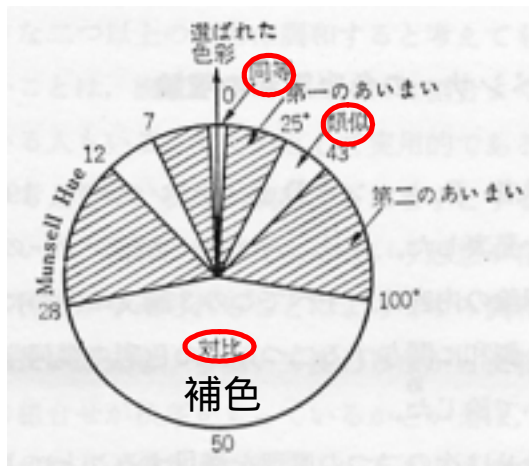


図1 二色の色相差と調和・不調和との関係 (参考文献[1])

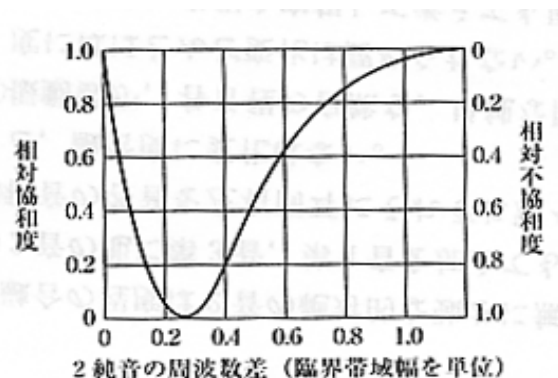


図2 周波数差と協和・不協和との関係分布 (参考文献[2])

柴田ら[3,4]は、協和・不協和音に画像を付加したときの調和・不調和と画像コントラストとの関係に、周波数差と協和・不協和との関係と同様な関係があると仮定し、画像と音との調和・不調和の分析を行った。図3は横軸を色相(0°から360°)、縦軸を画像コントラスト値にした分布を示す。図2と同じようなV字型の分布になっていることが分かる。自然画像の調和・不調和、二音和音の協和・不協和、両方の調和・不調和について心理実験を行った結果、協和・不協和音と画像の調和度は画像コントラ

ストによって変化することがわかった。

本研究は、画像コントラストモデルと周波数による二音和音の協和・不協和のモデルにより、画像と音の調和・不調和を自動的に推定するシステムを構築するために、図2と図3の分布の類似点に着目し、画像コントラストの仕組みを参考にして図2の二音和音の協和・不協和と周波数との関係のモデル化を行なう。ある画像・映像に調和する効果音・音楽をデータベースから検索を行うことや、逆に音楽に合う画像・映像を検索することも可能になる。また、作成されたコンテンツがお互い調和しているかを判定することも可能になり、編集などの時間を短縮することが可能になる。

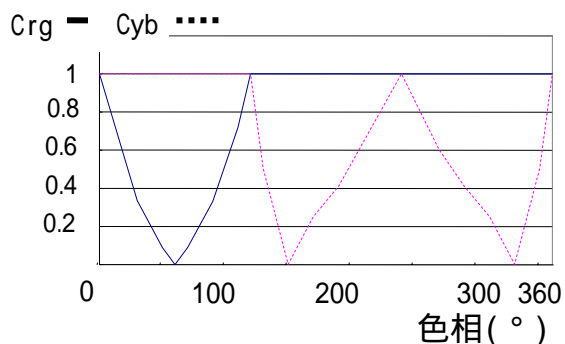


図3 色相とコントラスト値の関係分布

## 2. 画像コントラストのモデル

二音和音の協和・不協和の周波数による数理モデルを説明する前に、画像と音の調和・不調和と画像コントラストの関係について説明する。本研究の二音和音の協和・不協和の数理モデルは画像コントラストのモデルからヒントを得ている。

画像と音の調和・不調和度を変える画像側の要因として、心理実験の結果から(1)、(2)式のコントラストであることがわかった[3,4]。(1)、(2)式のr,g,bは画像全体のRGBの平均値とする。横軸を色相、縦軸をコントラスト値とした分布は図3のようになる(ただし、HSB色分布で明度=彩度=100とする)。色相範囲を限定すると、二音和音における周波数差と協和・不協和との関係分布(図2)に類似していることが分かる。

$$C_{rg} = \frac{|r - g|}{r + g} \quad (1)$$

$$C_{yb} = \frac{|r + g - 2b|}{r + g - 2b} \quad (2)$$

コントラストの計算は、画像全体の RGB それぞれの平均値を求めた後、(1)、(2) 式に代入し、コントラスト値  $C_{rg}, C_{yb}$  を求める。

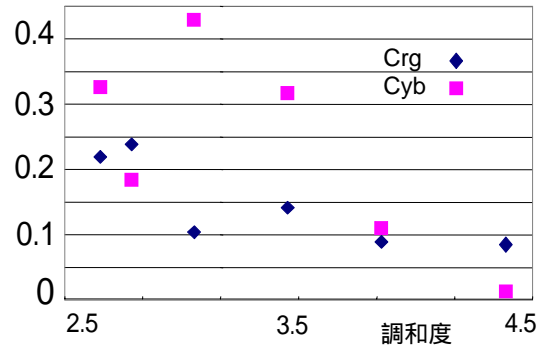
図 4 に協和・不協和音別に画像と音の調和とコントラスト値の関係を示す[3]。協和音については、コントラスト値  $C_{rg}$  は調和度と正の相関(図 4 a では負だが、調和度が小さいと調和となるため)があり、コントラスト値  $C_{rg}$  が大きいほど調和度が上がる傾向がある。一方、 $C_{yb}$  とは相関が弱い。不協和音については、コントラスト値  $C_{yb}$  と負の相関があり、コントラスト値  $C_{yb}$  が小さいほど調和度が上がる傾向がある。一方、 $C_{rg}$  とは相関が弱い(別の実験では  $C_{rg}$  と相関あり[4])。

つまり、調和度を上げる方法として、協和音にはコントラスト値  $C_{rg}$  が大きく、かつ、自然物が含まれる画像が好ましく、一方、不協和音にはコントラスト値  $C_{yb}$  (あるいは  $C_{rg}$ ) が小さい画像あるいは不調和と思われる画像が有効であることがわかった[4]。

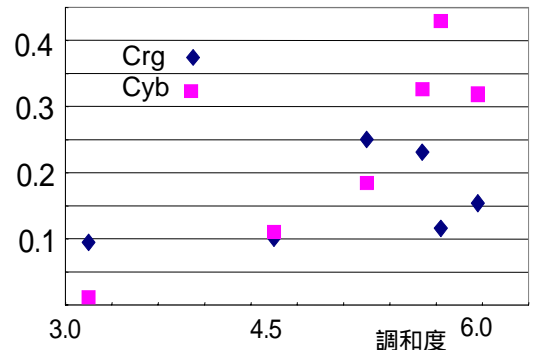
### 3. 二音和音の協和・不協和の数理モデル

前章では、画像と音の調和・不調和については二音和音が協和あるいは不協和かによって調和する画像は異なることを述べた。画像と音の調和・不調和を推定するためには二音和音が協和か不協和かを推定した上で画像を決める必要があることが分かる。そこで、Plomp & Levelt による二音和音の周波数の差と協和・不協和の分布を参考にして推定モデルを数式で構築する。モデルを構築する上で下記の基本方針を前提とする。

1. 1 オクターブ内でのモデル
2. 1 オクターブ内の最小周波数  $f_0$  と最大周波数  $f_1 (=2f_0)$  を利用



(a) 協和音



(b) 不協和音

図 4  $C_{rg}, C_{yb}$  と調和度との関係

3. ウェーバーの法則を利用
4. 画像のコントラストモデルを参考

方針 1,2 より、

- $f_1$ : 1 オクターブ内の任意の周波数
- $f_0$ : 1 オクターブ内の最小周波数=基準周波数
- $f_1 (=2f_0)$ : 1 オクターブ内の最大周波数

とする。(1) 式の画像コントラストモデルにおいて変数は  $r, g$  と 2 変数からなっており、二音和音では周波数と 1 変数である。図 2, 3 のような V 字型の分布にするため、方針 3 のウェーバーの法則を用い、2 つの変数  $f_a, f_b$  ((3), (4) 式) を考案した。 $f_a$  は周波数  $f$  と基準周波数  $f_0$  との差を基準周波数  $f_0$  で割り、 $f_b$  は最大周波数  $f_1$  と周波数  $f$  との差を基準周波数  $f_0$  で割った値とする。

$$f_a = \frac{f - f_0}{f_0} \dots (3)$$

$$f_b = \frac{f_1 - f}{f_0} = \frac{2f_0 - f}{f_0} \dots (4)$$

方針 4 より , (1), (2) 式の画像コントラストモデルを参考にして , 周波数コントラストモデル Cf' を (5) 式のように構築する .

$$C_f' = \frac{|f_a - f_b|}{f_a + f_b} \dots (5)$$

しかし , Plomp & Levelt の分布より , 最不協和になるとき Cf = 0 にならないので , (6) 式のようにパラメータ  $\alpha$  を導入した . これによって Cf = 0 のときの周波数の値が変化する .

$$C_f = \frac{|\alpha f_a - f_b|}{\alpha f_a + f_b} \dots (6)$$

(3), (4) 式を (6) 式に代入すると ,

$$\begin{aligned} C_f &= \frac{|\alpha(f - f_0) - (f_1 - f)|}{\alpha(f - f_0) + (f_1 - f)} \\ &= \frac{|(\alpha + 1)f - (\alpha + 2)f_0|}{(\alpha - 1)f - (\alpha - 2)f_0} \dots (7) \end{aligned}$$

Cf = 0 のとき , f の値を f' とすると ,

$$f' = \frac{\alpha + 2}{\alpha + 1} f_0 \dots (8)$$

ここで Zwicker らの臨界帯域幅 fw [5, 6] のモデル式を ,

$$f_w = \begin{cases} 100(f_0 \leq 500\text{Hz}) \\ 0.2f_0(f_0 \geq 500\text{Hz}) \end{cases} \dots (9)$$

と近似し , Plomp & Levelt の分布は  $\frac{f - f_0}{f_w} = 0.25$  のとき , 0 (= Cf) となるので (9) 式より ,

$$f' = \begin{cases} f_0 + 25(f_0 \leq 500\text{Hz}) \\ \frac{21}{20}f_0(f_0 \geq 500\text{Hz}) \end{cases} \dots (10)$$

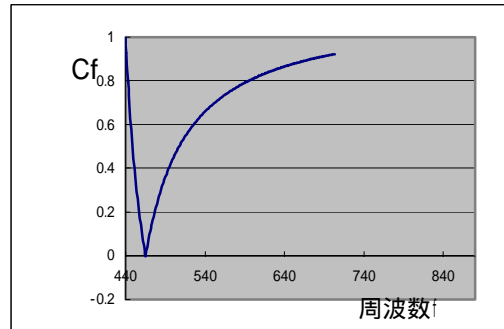
(8), (10) 式より , f' を消去し ,

$$\alpha = \begin{cases} \frac{f_0 - 25}{25}(f_0 \leq 500\text{Hz}) \\ 19(f_0 \geq 500\text{Hz}) \end{cases} \dots (11)$$

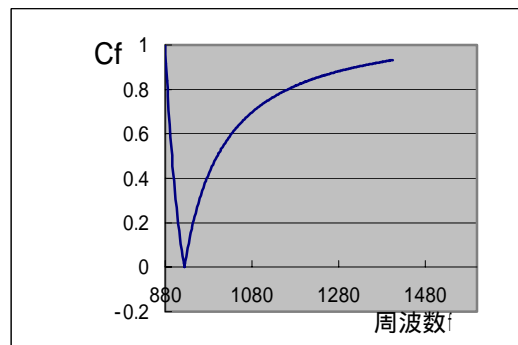
(7) 式に (11) 式を代入すると ,

$$C_f = \begin{cases} \frac{|f - (f_0 + 25)|}{\left(1 - \frac{50}{f_0}\right)f - f_0 + 75} (f_0 \leq 500\text{Hz}) \\ \frac{|20f - 21f_0|}{18f - 17f_0} (f_0 \geq 500\text{Hz}) \end{cases} \dots (12)$$

となる . 図 5 は , 基準周波数が 440Hz , 880Hz の時の周波数 f と Cf 値の分布を示す .



(a) 基準周波数 f0=440Hz



(b) 基準周波数 f0=880Hz

図 5 周波数 f と周波数コントラスト Cf との分布

図 2 の分布と多少異なるが , コントラストモ

デルから近似することが可能であることが分かる。

#### 4. まとめ

Plomp & Levelt による二音和音の協和・不協和と周波数差との関係分布を画像コントラストの仕組みを参考にしてモデル化を行った結果、同じような分布になることが分かった。調和・不調和に関しては、視覚あるいは聴覚情報処理に似たメカニズムがある可能性がある。

今後の課題として、画像コントラストモデルと二音和音の協和・不協和の周波数によるモデルによって、画像と音の調和・不調和を自動的に推定するシステムを構築し、評価を行う。将来的には、多くの画像や音データから調和するデータを抽出することが可能になり、編集などの時間が短縮できるシステムを構築するとともに、出来上がったコンテンツを評価するシステムにも応用していきたい。

#### 参考文献

- [1] 乾敏郎：“建築の色彩設計”，鹿島出版会，東京(1976)
- [2] Plomp and Levelt：“Tonal Consonance and Critical Bandwidth”，J. of Acoust. Soc. Am., pp.548-560 (1965)
- [3] 柴田滝也，川崎由加里，川崎舞佳：“画像と音の調和に関する心理的な分析とそのモデル化-協和・不協和音と画像コントラストとの関係-”，映像情報メディア学会技術報告(AIT2004-104)，vol.28，No.27，pp.41-44 (2004)
- [4] 柴田滝也，金森淳香：“画像と音の調和に関する心理的な分析とそのモデル化-協和・不協和音と画像コントラストとの関係(2)-”，画像電子学会技術報告(2004)(投稿中)
- [5] Zwicker, E., Flottorp, G., and Stevens, S. S. "Critical band width in loudness summation." *J. Acoust. Soc. Am.* 29, 548-557 (1957).
- [6] Zwicker, E., Fastl, H. *Psychoacoustics: Facts and Models*, Springer-Verlag (1990).