

眼球運動および瞳孔径計測に基づく音の快不快の推定

Decoding Perceptual Preference for Sounds from Eye Movement and Pupil Size Change

米家 惇^{*1 *2} リャオ シンイ^{*1} 古川 茂人^{*1} 柏野 牧夫^{*1 *2}
 Makoto Yoneya Hsin-I Liao Shigeto Furukawa Makio Kashino

^{*1} 日本電信電話株式会社 NTTコミュニケーション科学基礎研究所
 NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation

^{*2} 東京工業大学 大学院総合理工学系研究科
 Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

We propose a novel approach to decode perceptual preference for sounds from eye measurement. In the experiments, we measured eye movement and pupil size change while participants were presented with pairs of sounds (consonance and dissonance) and were asked to judge which sound was perceived as more annoying. Using new feature extraction methods for microsaccades, our approach could reach a classification accuracy of above 70%.

1. 背景

眼球運動や瞳孔は、視覚入力を制御する機能を持つ一方で、聴覚に関連した応答を示すことが知られる。たとえば、マイクロサッカド(跳躍性の微小眼球運動)は、聴覚的注意の誘発に伴って発生頻度が一時的に減少することが報告されている。

我々はこれまで、瞳孔径やマイクロサッカドの動特性が音の顕著性に関連した反応を示すことを明らかにしてきた。さらに、マイクロサッカドや瞳孔径変化に対して制御論モデルに基づく特徴抽出法を適用することで、眼球計測信号から「楽曲の好み」が一定精度で推定できることを示してきた。しかし、楽曲の好みの推定にあたっては、眼球の特徴以外に音響信号の特徴などの複合的な情報を考慮する必要があった。

そこで本発表では、音響信号の特徴量を推定の手がかりとすることなく、眼球計測に基づく特徴量のみを用いた場合の推定性能を検証するため、単一の音に対して知覚される「快不快」の推定を試みる。具体的には、西洋音楽において一般に美しいとされる「協和音」と、そうでない「不協和音」を聴取した際の眼球計測を行い、眼球特徴量に基づいた快不快の推定を行う。

2. 実験

20-40代の健常成人22名が実験に参加した。実験参加者は、2つの音の組を繰り返し聴取し、1音目(以下、音A)と2音目(以下、音B)を比較してより不快と感じる音を左右のボタン押しによって回答した(図1)。

刺激音には、表1に示す協和音3種類および不協和音3種類を用いた。音Aか音Bの一方は協和音、もう一方が不協和音となるよう設定した。各音の持続時間は4秒とし、各ペアにおいて音Bは音Aの終了直後に呈示を開始した。協和音と不協和音を一つずつ含む計9通りのペアを各8回ずつ用い、合計72試行をランダム順に呈示した。

実験参加者が音楽を聴取する際の眼球運動を、SR Research社製 EyeLink-CL Systemを用い、1000Hzのサンプリングで計測した。計測中、実験参加者には正面に設置されたモニター中央の注視点に視線を向けるよう教示した。モニターにはグレーの背景に黒の注視点のみを表示した。

表1. 実験に用いた刺激音の一覧

	音	音程名	半音数	純正律 周波数比
協和音	C4, F4	完全4度	5	3:4
	C4, G4	完全5度	7	2:3
	C4, C5	完全8度	12	1:2
不協和音	C4, C#4	短2度	1	15:16
	C4, D4	長2度	2	8:9
	C4, B4	長7度	11	8:15

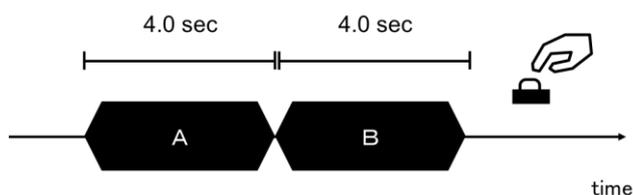


図1. 実験の概要図

3. 方法

実験参加者が1つのペアを聴取した8secのデータを1試行の単位とした。各試行から計算された多次元特徴ベクトルと、ボタン押しによる回答値(主観評定値)の対応関係を学習し、交差検定を用いて推定精度の評価を行った。

3.1 特徴抽出

計測データに基づいて、各試行において下記の方法により眼球運動および瞳孔径変化に関する特徴量を計算した。

(1) 眼球運動に関する特徴量

眼球運動に関する特徴量として、マイクロサッカドの発生頻度、振幅、最大速度を計算した。さらに、本研究ではマイクロサッカドの波形が二次系のステップ応答として記述できることに着目し、波形の振動性を表す減衰係数、波形の速応性を表す固有角振動数を特徴量として計算した。

マイクロサッカドの検出は、Engbertら(2003)の提案したアルゴリズムに基づいて行った。具体的には、水平方向の眼球位置の差分系列(速度)を計算した上で、速度の絶対値が閾値

(標準偏差の 8 倍)以上となった点を開始点として定義した。ただし、2°以上の振幅をもつものはラージサッカードであるとして除去した。ノイズによる誤検出を低減するため、瞬目の前後 25 ms 以内に検出された点は除去し、5 ms の範囲内で隣接する開始点群については最後の点以外を除去した。また、マイクロサッカードが両眼性であることから、左右の目でそれぞれ検出された開始時刻の差が 10msec を超えるもの、片眼でのみ検出された点は除外した。

(2) 瞳孔径変化に関する特徴量

瞳孔は自律神経系の支配を受けて驚きや快不快のような感情に応じて変化を示すことが知られる。本研究では瞳孔径変化の計測データの平滑化を行った上で、一定の基準に基づいて極小点/極大点を検出することにより散瞳(瞳孔の散大)/縮瞳(瞳孔の縮小)の開始点を抽出し、散瞳/縮瞳の発生頻度、振幅、平均速度、持続時間等を瞳孔径変化の特徴量として計算した。マイクロサッカードと同様、基本的に縮瞳や散瞳も両眼性で生じるイベントと考えられるため、左右の目で検出された開始時刻の差が 100msec を超えるもの、片眼でのみ検出された点は除外した。

以上によって計算された各特徴量について、音 A 呈示中の平均値が音 B 呈示中の平均値に比べて大きい場合には 1、そうでない場合には 0 として、特徴量毎にバイナリ化を行った。バイナリ化を行うことで、生体特徴量の個人内変動や外れ値による影響を低減する効果が期待できる。

なお、音 B の呈示終了後 2sec 以内に回答のなかった試行、音 A あるいは音 B の聴取時にマイクロサッカードが一つも検出されなかった試行は不良試行として排除した。この基準によって、全 1584 試行中 957 試行、被験者 22 名中 17 名が解析の対象となった。

3.2 学習と推定

被験者毎に重回帰分析による学習と推定を行い、Leave-one-out cross validation によって評価した。

学習データに基づき、前述の特徴量群を説明変数、被験者の回答値(音 A を不快と回答した場合:1, 音 B を不快と回答した場合:2)を目的変数とする重回帰分析を行った。このとき、VIF 統計量に基づいた特徴選択を行ったうえで、最適な特徴量の部分集合を SFSS 法によって求めた。

学習によって得られた重回帰係数に基づき、テストデータに対して求められた推定値が 1.5 未満の場合は「音 A」が選ばれたものとして推定、推定値が 1.5 以上の場合は「音 B」が選ばれたものとして推定した。

主観評定値の推定に対する比較として、目的変数を協和/不協和の刺激音特性そのもの(音 A が不協和音の場合:1, 音 B が不協和音の場合:2)とした場合の推定精度も同様の手法により求めた。

4. 結果

4.1 協和音/不協和音に対する主観評定

ボタン押しの回答によって不協和音が「不快」と判断された割合は、被験者平均で 89.2±12.6% (mean±S.D.)であった。17名の被験者のうち、100%の割合で不協和音を不快と判断した被験者は3名、80%以下の割合で判断した被験者は3名であった。ペア毎の結果では、(C4, C5)と(C4,C4#)のペアにおいて不協和音を不快と判断された割合が 97.0%と最も高かった(表2)。

表2: 音のペア毎の主観評定(不協和音を不快と回答した割合)

	C4, F4	C4, G4	C4, C5
C4, C#4	87.6%	95.4%	97.0%
C4, D4	89.9%	95.5%	87.2%
C4, B4	92.8%	86.9%	86.9%

表3: 音のペア毎の推定精度(主観評定の推定)

	C4, F4	C4, G4	C4, C5
C4, C#4	74.2%	64.4%	72.4%
C4, D4	67.7%	65.2%	68.4%
C4, B4	68.0%	83.3%	67.2%

表4: 音のペア毎の識別精度(協和/不協和の識別)

	C4, F4	C4, G4	C4, C5
C4, C#4	62.9%	59.8%	67.9%
C4, D4	65.7%	60.7%	66.9%
C4, B4	68.0%	75.0%	65.7%

4.2 眼球計測に基づく推定

重回帰分析によって推定された結果と、実際的主観評定値との一致率は 71.4±7.4%であり、chance level よりも有意に高い水準にあった(両側符号検定, $p<.001$)。ペア毎の結果では、(C4, G4)と(C4, B4)のペアにおいて推定精度が 83.3%と最も高かった(表3)。ペア毎の主観評定の推定精度と「不協和音を不快と回答した割合」(表2)との間には有意な相関は認められなかった($r=-0.43$, $p=0.25$)。重回帰分析において最も多く選択された特徴量はマイクロサッカードの発生頻度であり、17名の被験者のうち13名において有効な特徴量として選択された。

上記と同様に、重回帰分析によって主観評定ではなく協和/不協和の識別を行った結果、推定精度は 67.6±8.4%であり、主観評定の推定に比べ精度は低かった(表4)。

5. まとめと考察

眼球の特徴量のみに基づいて、協和音/不協和音に対して感じられる快不快の判断を一定精度で推定可能であった。また、刺激の特性(協和/不協和)の識別に比べ、主観評定の推定の精度が高かったことから、眼球の特徴量が単に刺激の性質を推定しているだけでなく、快不快の知覚そのものに関連していることが示唆された。

参考文献

- [Liao 2015] H. I. Liao, S. Kidani, M. Yoneya, M. Kashino, & S. Furukawa: Correspondences among pupillary dilation response, subjective salience of sounds, and loudness. *Psychonomic Bulletin & Review*, 2015.
- [Yoneya 2014] M. Yoneya, H. I. Liao, S. Kidani, S. Furukawa, & M. Kashino: Sounds in Sequence Modulate Dynamic Characteristics of Microsaccades. In 37th ARO MidWinter Meeting, 2014.
- [Engbert 2003] R. Engbert, R. Kliegl: Microsaccades uncover the orientation of covert attention, *Vision Research*, vol. 43, no. 9, pp. 1035–1045, 2003.