自律型ＡＩスピーカーに向けた
表情による感情推定を用いた音楽提示法

長久保　風雅
指導教員　林　実
明星大学　理工学部　総合理工学科　電気電子工学系　林研究室

近年注目されている AI スピーカーは、従来から指示型であった。そこで自らアクションを起こすことのできる「自律型」AI スピーカーの実現を目指し研究を進めている。ここでは、画像情報における人の表情に着目し、表情による感情推定から音楽を流すことのできるシステムへ向けて実験を行った。その結果、表情による感情推定値が、音楽感情相関を示す AVSM の尺度別に有意差が見られた。

キーワード：表情認識，感情推定，音楽提示，AVSM

1.　はじめに

近年、AI（人工知能）や IoT の発達によりさまざまな情報・モノがつながり、学習・認識を行い、多様な機能が生まれている。その中で、人とモノをつなげるインターフェースとして AI スピーカーが注目されている。

AI スピーカーは人が発した言葉を認識し、音楽を流すだけでなく、人に返答したり、他の機器を制御したりすることができる。しかし、スピーカーは基本的に指示を待つだけであって、自らが理解し、アクションを起こすことができない。そこで、このような指示型 AI スピーカーに対して、自ら場の状況や人の感情を認識し、アクションを起こすことのできる「自律型」AI スピーカーの実現を目指して、研究を進めている。

本稿では、自律型 AI スピーカーの実現を目指すために、画像中における表情に着目し、スピーカーが画像情報中の表情から感情を推定し、その感情推定に合う音楽を流すことができるシステムへ向けて実験を行なったので報告する。

2.　表情認識からの感情推定について

表情から感情推定する手法・ツールは複数存在する。今回は Microsoft 社の提供する Microsoft Azure Cognitive Services 内の Face API を利用した。この Face API には Microsoft Research で研究開発された「複数の深層ネットワーク学習を用いた画像に基づく表情認識技術」[1]など Microsoft 社が蓄積してきた技術が使われている。

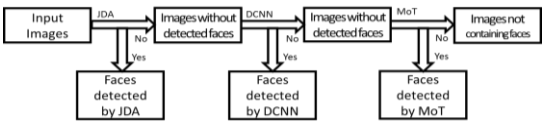

図 1 複数の DNN を用いた画像に基づく表情認識技術

図 1 に複数の DNN を用いた表情認識技術のシステムの流れを示す。表 1 に本実験で用いた Face API の出力する感情推定値の詳細を示す。

表 1 Face API による感情推定値の詳細

出力値	0.0～1.0 の信頼度
感情の個数	8 個
感情の種類	neutral（中立）、happiness（喜び）、surprise（驚き）、contempt（軽蔑）、anger（怒り）、fear（恐怖）、disgust（嫌悪感）、sadness（悲しみ）
出力方法	8 個の感情価の信頼度が割合で出力される。（8 個の合計が 1.0）

3.　音楽と感情との関係について

音楽と感情との関係については音楽心理学の領域で現在も研究が進んでいる。そのなかでもよく心理学実験等で用いられるのが、表 2 に示す谷口氏の 6 尺度２４形容詞項目で表す音楽作品の感情価測定尺度（Affective Value Scale of Music ; AVSM）である[2]。今回、実験における音楽作品の感情価数値の引用に、この AVSM を用いた。尚、表 2 に谷口氏が後に行った AVSM の測定において各尺度の高得点を得た作品を付記する。

表 2　AVSM 各尺度別形容詞と高得点を得た作品

尺度	形容詞	作品名（作曲者）
抑鬱	沈んだ、哀れな、悲しい、暗い	弦楽とオルガンのためのアダージョ　ト短調（アルビノーニ）
高揚	明るい、楽しい、うれしい、陽気な	シンフォニア 変ロ長調「シバの女王の入城」（ヘンデル）
親和	優しい、いとしい、恋しい、おだやかな	G 線上のアニア（J.S.バッハ）
強さ	強い、猛烈な、刺激的な、断固とした	エチュード 10-12「革命」（ショパン）
軽さ	きまぐれな、浮かれた、軽い、落ち着きのない	プリンク・プランク・プランク！（アンダーソン）
荘重	厳粛な、おごそかな、崇高な、気高い	弦楽とオルガンのためのアダージョ　ト短調（アルビノーニ）

4.　実験

AVSM の感情価と表情認識による感情推定との関係を調べるために、以下の実験を行った。

4.1.　実験方法

まず下記に示す２つの実験方法で、使用する画像データを収集した。

4.1.1.　実際に撮影した表情画像を用いた方法(実験①)

図 2 に実験環境を示す。以下の手続きで、被験者の表情を撮影した。

・被験者を固定されたイスに座ってもらい安定状態にしたうえ、カメラで表情をビデオ撮影した。

・音楽に起因した表情が作れるようにするため、表 2 に示す AVSM の各尺度にて高得点を得た音楽をスピーカーから流し、約 1 分間被験者に聴取してもらった。音楽はその表情をつくり終わるまで継続して流した。

・モニターに形容詞を表示し、その形容詞の表情をつくってもらった。（約５秒継続）
・撮影したビデオのなかから各表情の安定したところ（変化がないところ）をキャプチャーし、画像データとして保存した。

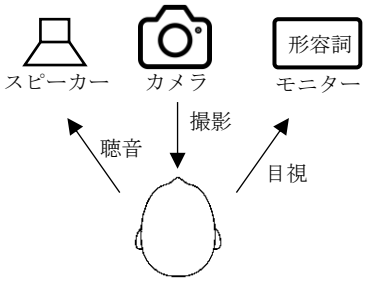

図 2 表情画像の撮影実験環境

4.1.2. 画像検索による表情画像を用いた方法(実験②)

AVSM を構成している形容詞２４項目別の表情を、インターネット画像検索を用いて収集し、ダウンロードした。表 3 に収集検索条件を示す。

表 3 収集検索条件

使用した画像検索エンジン	Google 画像検索 https://www.google.co.jp/imghp?hl=ja
検索方法	「●●表情」または「●●顔」で検索（例・「沈んだ」⇒「沈んだ表情」「沈んだ顔」）
収集方法	検索結果の上位※画像より「単独」「人間」「写真」であるものを選択し、さらに Face API にて予備の顔認識を実施し　正しく表情認識できるものを収集した。　※2018 年 9 月 20 日(木)時点

4.1.3. 画像情報中の表情認識とデータ解析

表 4 に収集した画像情報の条件を示す。収集した画像を Python で実装した Face API を用いて表情認識させ、感情推定値を記録した。この記録から AVSM の各尺度別に平均値をとり、実験結果としてまとめた。

表 4 実験に用いた画像データの条件

収集方法	実際に表情を撮影	画像検索を用いて収集
枚数	192 枚	120 枚
被験者数	8 名	－
アングル	固定（正面のみ）	変動（正面、斜め　等）
表情の作り方	人為的	自然

4.2. 実験結果

今回は AVSM 尺度別の Face API 感情推定値の結果の平均をレーダーチャートにまとめ、それぞれ図 3 から図 8 までに示す。図中の「実験①、実線」は画像収集の際に実際に撮影した表情画像を用いた結果を、「実験②、点線」は画像検索による表情画像を用いた結果を示す。

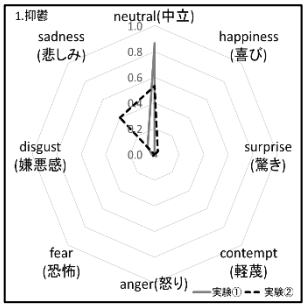

図 3「抑鬱」の実験結果

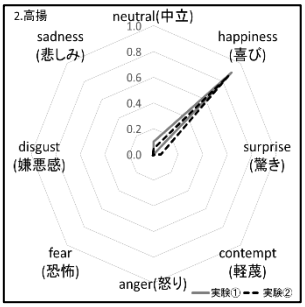

図 4「高揚」の実験結果

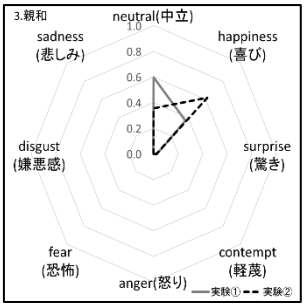

図 5「親和」の実験結果

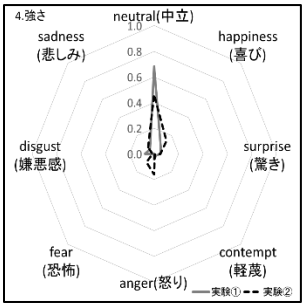

図 6「強さ」の実験結果

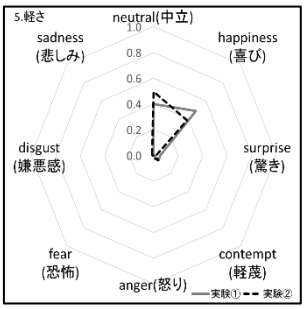

図 7「軽さ」の実験結果

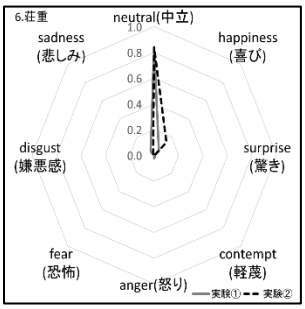

図 8「荘重」の実験結果

5. 考察

今回の実験では、AVSM の尺度別形容詞と Face API の表情認識による感情推定値にどのような関係があるかを測定した。６つの尺度を比較すると有意差が見られる部分があった。特に“高揚”のように特徴が強く表れたものもあり、強く特徴の出る尺度に関しては、表情による感情推定から音楽を流すことが可能ではないかと考えられる。

一方、“強さ”“荘重”のように「neutral(中立)」成分が強く特徴がつかめない尺度も存在する。また今回２種類の方法でサンプルを集めたが、“抑鬱”を除く５つの尺度で同等の特徴的な結果を求めることはできたが“抑鬱”だけがなぜ２種の方法で結果が異なるか不明確であった。これら結果は今後検討の余地があると考える。

今回は表情にのみに着目したが、表情以外の方法による感情推定（画像中の色、人のしぐさや、映像における音声など）も合わせ、研究を深めていきたい。

6. まとめ

本研究では、自律型 AI スピーカーを目指し、画像情報の表情から感情を推定し、その感情推定に合う音楽を流すための実験として、AVSM の尺度別形容詞と Face API の感情推定値の関係を求めた。その結果、特徴の強弱はあるものの尺度別に有意差が見られた。本手法は、自律型 AI スピーカーにおける状況や感情認識へのアプローチになると思われる。

参考文献
[1] Yu, Zhiding, Cha Zhang. "Image based static facial expression recognition with multiple deep network learning." Proceedings of the 2015 ACM Int Confer Multi Inter ACM：435–442, 2015.
[2] 谷口高士. "音楽作品の感情価測定尺度の作成および多面的感情状態尺度との関連の検討." 心理学研究 65.6：463-470, 1995.