

顔と声を同時提示した際の主観年齢分析

及川 隼平[†]

群馬工業高等専門学校
専攻科 生産システム工学専攻[†]

川本 真一[‡]

群馬工業高等専門学校
電子情報工学科[‡]

1 はじめに

日常生活において、「あの人は顔と声の印象が一致しない」と感じることもある。マガー効果 [1] や腹話術効果 [2] のような現象は、視覚情報と聴覚情報が相互に影響することを強く示唆している。

視聴覚メディアをデザインするには、顔と声を組み合わせた際の主観的な印象を調べる必要がある。しかし、顔や声単体の知覚に関する研究はあるものの、顔と声を組み合わせた研究は少ない [3]。そこで本研究では、数値として扱うことができる年齢推定に着目し、顔と音声の組み合わせによる印象の変化を分析することを目的とした。

先行研究では、顔の静止画と声とを組み合わせたときの主観年齢の変化を分析した [4]。本研究では、顔の情報を増やした場合の主観年齢の変化の分析を行う。

2 実験内容・結果

2.1 実験条件

モニタは MacBook Pro (Apple) の内蔵ディスプレイを、ヘッドホンは MDR-CD900ST (SONY) を用いた。刺激の再提示は認めなかった。

顔画像は、The Japanese Female Facial Expression (JAFPE) Database [5] を用いた。

音声は、北陸先端科学技術大学院大学 研究用日

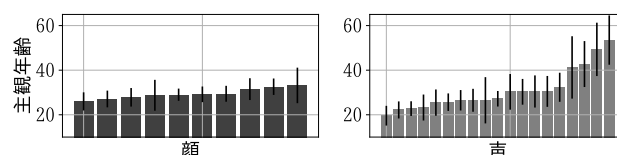


図1 顔の静止画または声単体を提示した場合の主観年齢

本語感情音声データベース (JAIST-ESD) [6] と株式会社 ATR-Promotions の ATR キャラクター音声データベース I を用いた。

2.2 顔の静止画または声単体を提示した場合

2.1 節で述べたデータについて、被験者 13 名を対象に主観年齢を評価する実験を行った。顔画像は 30 枚を時間制限なしで提示した。音声は 128 個を再提示なしで提示した。実験結果を図 1 に示す。ここで、エラーバーは各顔または声の標準偏差を示している。顔画像は 26.0~33.2 歳、音声は 19.6~53.5 歳の主観年齢のデータであることが分かった。

2.3 リップシンクアニメーションを付与した顔と声を同時に提示した場合の主観年齢

先行研究では、顔の静止画と声では声優位であることが分かった [4]。音声は時間変化するが、顔の静止画は時間変化しないため、顔の情報よりも声の情報の方が情報量が多いと考えられる。そこで、顔にリップシンクアニメーションを付与することによって情報量を増やすとともに、実際の会話に近づけることができると考えた。リップシンクアニメーションの作成には Galatea [7] を用いた。被験者 10 名を対象に動画 20 本を再提示なしで提示し、主観年齢を評価する実験を行った。結果を図 2 に示す。顔と声を同時に提示した場合を基準としたとき、顔との相対誤差は 0.0589、声との相対誤差は 0.0852 となり、顔優位となった。

Analysis of Subjective Age by Combination of Faces and Voices

[†] Shumpei Oikawa, Advanced Production Systems Engineering Course, National Institute of Technology, Gunma College

[‡] Shinichi Kawamoto, Department of Information and Computer Engineering, National Institute of Technology, Gunma College



図2 リップシンクアニメーションを付与した顔と声を同時に提示した場合の主観年齢の相対誤差

3 考察

3.1 顔と声を同時に提示した場合の主観年齢

先行研究では、顔の静止画と声を同時に提示した場合、声優位であった[4]。これに対し、2.3節の実験によって顔にリップシンクアニメーションを付与することによって、顔優位となった。これらの結果は、顔の情報量によって重要度が変化することを示唆している。

3.2 平均顔と音声モーフィングによる主観年齢の制御

2.2節のデータを用いて、平均顔とモーフィング音声を作成し、主観年齢がどのような関係になるか調べた。

平均顔

2つの顔を合成して作成したコラージュ画像を平均顔と呼ぶ。顔の印象の加工方法として平均顔を取り上げ、主観年齢を付与し、元の顔との主観年齢の関係を調べた。平均顔の生成にはOpenCV[8]とDlib[9]を用いた。被験者6名を対象に顔画像45枚を時間制限なしで提示し、主観年齢を評価する実験を行った。平均顔の主観年齢の分布を表1に示す。2つの刺激の中間の関係である内挿だけではなく外挿になる場合も多く見られ、主観年齢が26.0～33.2歳であった元データを23.3～35.5歳に広がった。平均顔は輪郭がぼやけ、シワがなくなるため、顔から読み取られる情報が減り、年齢幅が広がったと考えられる。

モーフィング音声

2つの音声の中間の声を生成することを音声モーフィングと呼ぶ。声の印象の加工方法として音声モーフィングを取り上げ、主観年齢を付与し、元の音声との主観年齢の関係を調べた。モーフィング音声の生成にはスペクトルモーフィング[10]を用

表1 平均顔とモーフィング音声の主観年齢のデータ分布

	平均顔	モーフィング音声	
		JAIST-ESD	ATR
外挿 (上側)	15	9	14
内挿	16	9	20
外挿 (下側)	13	12	38

いた。被験者6名を対象に音声74個を再提示なしで提示し、主観年齢を評価する実験を行った。モーフィング音声の主観年齢の分布を表1に示す。主観年齢が19.6～53.5歳であった元データを12.0～57.2歳に広がった。また、内挿よりも外挿になるケースが多く、特に下側の外挿になるケースが多いことが分かった。これは、モーフィングによって明瞭度が下がることが影響していると考えられる。

以上の通り、平均顔やモーフィング音声の印象は必ずしも平均にならないということが分かった。今後はこれらのルールについて検討し、制御規則について分析を行う必要がある。

4 まとめ

顔と声を同時に提示した場合の主観年齢の変化を分析した。顔の静止画と声では声の方が影響が強くなり、顔のリップシンクアニメーションと声では、顔の方が影響が強くなることが分かった。

謝辞

本研究はJSPS 科研費JP25240026, JP15K21024の助成を受けたものです。

参考文献

- [1] McGurk H., MacDonald J., Nature 264 (5588), 746-748, (1976)
- [2] Bertelson P., Aschersleben G., Psychonomic Bulletin and Review, 5, 482-489, (1998)
- [3] Moyse E., Psychologica Belgica, 54 (3), 255-265, (2014)
- [4] S. Oikawa, S. Kawamoto, 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE), 737-738 (2018)
- [5] Michael J. L. et al., <<http://www.kasrl.org/jaffe.html>> (2019-1-5)
- [6] 赤木 正人 他, "北陸先端科学技術大学院大学 研究用日本語感情音声データベース (JAIST-ESD)," (2016)
- [7] 嵯峨山 茂樹 他, 情報処理学会研究報告音声言語情報処理 (SLP), 2002-SLP-045, 57-64, (2003)
- [8] OpenCV team, <<https://opencv.org/>> (2019-1-6)
- [9] Davis E. King, <<http://dlib.net/>> (2019-1-6)
- [10] 滝澤 照太, 川本 真一, 平成 29 年度北陸地区学生による研究発表会講演論文集, p.141 (2018)