

音声を用いた主観年齢の推定

－顔画像との比較－

Subjective Age Estimation using Speech Sounds

- Comparison with Facial Images -

○西本真由香¹ 宮本直幸¹ 藤澤隆史¹ 長田典子¹

(1 関西学院大学理工学部情報科学科)

E-mail: m_nishimoto@ksc.kwansei.ac.jp

1. 緒言

自己がイメージする自分の年齢を主観年齢と定義し研究を行ってきた[1]。顔画像を用いた主観年齢は総じて若年視の傾向にあることが確認されている。

本研究では、コミュニケーションにおける発話の要素を考慮し、音声を用いた主観年齢の推定手法を提案する。得られたデータに非線型回帰分析を用いて主観年齢値を算出し[2]、顔画像を用いた主観年齢値との比較検討を行う。

2. 音声を用いた主観年齢推定実験

音声データベース：現在20歳から66歳までの男女56名分の音声データを実年齢がほぼ均等数になるように収録している。音声データは、エレクトレットコンデンサーマイクロホン（SONY製ECM-MS957）とPCを用いて、日本語5母音を録音したものである。

音声評定手順：年齢・性別の異なる評定者に対して、音声データベースから評定者の実年齢±10歳に属する男女それぞれ15名分の音声を選択する。2性別×15名計30個のデータを刺激として評定課題を行う。評定者はランダム順に呈示された音声データに対して「自分より年上か年下か」を評価する。評定者には選択音声に対して「自分より年上か年下か」の5段階評定をもとめた。

主観年齢推定手順：X軸に実年齢差（＝音声の実年齢－評定者の実年齢）、Y軸に音声に対する評定結果をとると、右上がりの分布となる（図1(a)）。この分布の近似曲線のX切片が評定者群の主観年齢シフト値を表す。分布からロジスティック関数によって近似できると仮定し、各評定者群の分布に対して非線型単回帰分析を適用する。ここで扱うロジスティック関数は以下の式を用いる。

$$y = \frac{4}{1 + \exp(-a(x - b))} - 2$$

パラメータaは曲線の傾き、bはx軸との交点を表す。非線型単回帰分析によりa、bの値を推定し、bを評定者の主観年齢シフト値と定義する。

3. 結果および考察

プレ実験として25歳から36歳の評定者11名に対して音声評定課題を実施し、得られたデータに対して主観年齢シフト値を算出した。重決定係数の値が低いデータ（ $R^2 < 0.000$ ）を除外し、10名分のデータセットに対して平均と標準偏差を求めた。結果を表1および図1に示す。

表1. 主観年齢シフト値についての記述統計量

	顔画像	音声
平均	-1.75	5.738
標準偏差	2.66	9.713

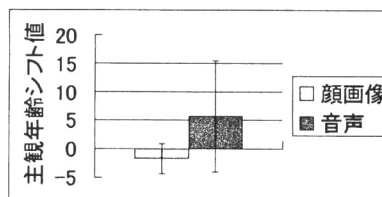


図1. 音声・顔画像別の主観年齢シフト値の比較

音声による主観年齢は、顔画像によるものとは異なり、正の値をとっている。また推定値のばらつきが顔画像によるものと比較して極めて大きい。主観年齢推定においては、音声情報は顔画像情報とは異質な意味合いを持つ可能性を示している。

4. 結言

音声を用いた主観年齢の推定手法を提案し、顔画像による主観年齢との比較検討を行った。今後の課題は、性別や年齢層別など評定者のプロフィールの違いによる主観年齢の差異を考察することである。

参考文献

- [1] 宮本, 陣内, 藤澤, 長田, 井口: “顔画像を用いた自己の主観年齢の推定”, 電子情報通信学会論文誌A, J90-A(3), 240-247, (2006).
- [2] 藤澤, 宮本, 長田, 井口: “顔画像を用いた自己の主観年齢の推定 –若年視傾向の規定要因に関する考察–”, 日本顔学会誌, 7(1).