

音声データを対象とした声質の特徴量分析による 印象推定方式

青木 絢太[†] 岡田 龍太郎[†] 中西 崇文[†]

武蔵野大学 データサイエンス学部 データサイエンス学部[†]

1. はじめに

人間はコミュニケーションにおいて、言語情報のみならず声質などの非言語情報も、印象を表現する上で重要な役割を持っている。一般的に、声質は、その人の印象や話している内容の印象を想起させるのに重要な要因である。相手に伝えたい内容に応じて声質を変えることを積極的に行うことにより、感情をうまく相手に伝え、コミュニケーションを円滑に進めることが可能となると考えられる。

ユーザの声の音声データを入力し、その音声データから音声特徴を抽出し、その特徴に基づいて想起されうる印象を推定することが出来れば、自身の声質の印象を客観的に把握でき、より良いコミュニケーションを実現する一助になると考えられる。

本稿では、音声データを対象とした声質の特徴量分析による印象推定方式について示す。本方式は、ユーザの音声データを入力とし、音声の声質が聞き手に与える印象を出力する。本方式では、あらかじめある発話集団の音声データに対して印象のタグ付けを行い、音声の特徴量のクラスタリングによって印象タグと特徴量を結びつける。利用時には、ユーザが発話した音声データから音声特徴量を抽出し、声質と印象をタグ付けしたデータベースを参照することでユーザの入力した音声の印象を出力する。

本方式の実現によって、自身の声の特徴を客観的に知ることが可能となり、相手に伝えたい印象を表現できているか確認することができる。

2. 関連研究

既存の研究として、池本らの感情表出時の声質評価尺度の作成[1]では感情を表出した際の表現語を抽出し、聴取実験によってその因子構造

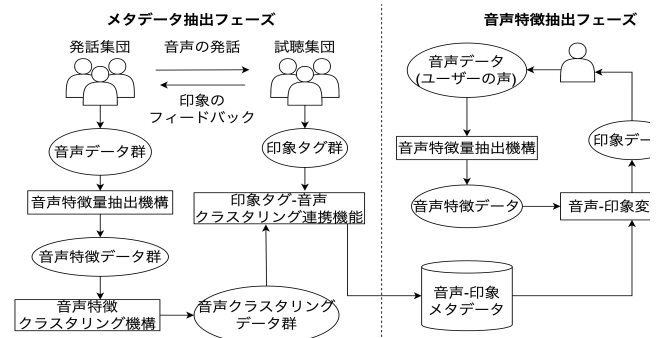


図1：特徴量クラスタリング方式を用いた音声印象自動抽出方式の全体像

を基本周波数とともに調べるといふものがある。本稿では、感情ではなく声質表現語そのものを複数の音声特徴量によるクラスタリングによって特定する。

3. 特徴量クラスタリングを用いた音声印象自動抽出方式

3.1 音声印象自動抽出方式の全体像

図1に、音声印象自動抽出方式の概要を示す。本方式は大きく分けて、メタデータ抽出フェーズと音声特徴抽出フェーズからなる。メタデータ抽出フェーズでは、サンプルとなる音声データを560個収集し、それぞれの音声データに木戸らの通常発話の声質に関連した日常表現語の抽出[2]に使われている印象語25語と独自に印象語18語を追加し、表1のように計43語としてそのうち複数を付与した。それらの音声データの音声特徴量を抽出し、クラスタリングすることで、そのクラスタに含まれる音声データに付与した印象語の多いものをそのクラスタの印象語として保持する。音声特徴抽出フェーズでは、ユーザ自身が自分の声を収録した音声データを入力し、メタデータ抽出フェーズで求めたクラスタのどれに近しいか計算をし、そのクラスタの印象語をその音声の印象として付与する。

3.2 音声印象特徴量抽出とクラスタリング

各音声データから音声の特徴を表すゼロクロス

A Method for Estimating Impressions Based on Feature Analysis of Voice Characteristics for Voice Data
Kenta Aoki[†], Ryotaro Okada[†], Takafumi Nakanishi[†]
[†] Musashino University, Department of Data Science

表 1：付与した印象語の一覧

かわいい声	クールな声	優しい声	明るい声
澄んだ声	色っぽい声	機械的な声	張りのある声
通る声	鼻の詰まった声	かすれた声	がらがら声
だみ声	くもった声	潰れた声	ドスの効いた声
恐怖を感じる声	迫力を感じる声	気分が良い声	気分が悪い声
子供っぽい声	若い声	中年の声	老人の声
渋い声	男性的な声	女性的な声	中性的な声
金切り声	高い声	弱々しい声	ハキハキした声
生き生きした声	響きのある声	感情の薄い声	落ち着きのある声
落ち着きのない声	品のある声	裏のありそうな声	疑問のある声
舌足らずな声	太い声	遅い声	

率, MFCC (20 次元), クロマシフト, スペクトラルセントロイド, スペクトラルバンド幅, の 25 次元の特徴量を抽出する。

抽出した 25 次元の特徴量ベクトルを k-means 法でクラスタリングする。

3.3 印象タグ-音声クラスタリング連携

事前に手動で印象語を複数付与した印象語について, 3.2 節クラスタリング結果に基づき重みを計算することとする。同じクラスタ内の音声データに多数付与されている印象語は, そのクラスタの印象を表す印象語となる。しかし, クラスタ内の音声データに多数付与されている印象語であったとしても, 他のクラスタ内でも同様に付与されている印象語は, そのクラスタの印象を表すとは言い難い。この関係は, 文書検索でよく用いられる tf-idf の考え方に通じる。

本稿では, この tf-idf の計算式を用いて得られた, クラスタ内の音声データに付与された印象語のクラスタごとの出現頻度と回数による計算結果から, クラスタの傾向を表す印象語を決定する。これによってクラスタ内の音声データがどのような印象語もつ傾向にあるかを推定することができる。

3.4 クラスタ中心との距離計算

新たに入力した音声データを対象として, 3.2 節で決定したクラスタとの距離計算を行い, 一番近かったクラスタの 3.3 節で決定した印象語を, 入力された音声データの印象語として出力する。

4. 実験

実際に新規に印象語を付与し, 入力した音声データがどのような結果を得たかを表 2 に示す。

表 2：新規データを入力した際に提示された印象語と事前に付与した印象語

音声データ	事前に付与した印象語	提示された印象語
Test1. mp3	男性的, 感情が薄い, 落ち着きがある	男性的, 中年, 感情が薄い
Test2. mp3	女性的, かわいい, 生き生き, ハキハキ	生き生き, かわいい, 女性的, ハキハキ
Test3. mp3	男性的, 遅い	女性的, 遅い
Test4. mp3	女性的, クール, 中年, 色っぽい	女性的, クール, 感情が薄い,
Test5. mp3	かわいい, 女性的, 子供っぽい	子供っぽい, かわいい

提示された印象は各クラスタの tf-idf の値が 0.3 以上のものとした。

表 2 のように事前に主観で付与した印象語とシステムが提示した印象語が合致していることが多いと考えられる。

5. おわりに

本稿では音声データを対象とした特徴量抽出, およびクラスタリングによる音声データの印象を表す印象語設定と新規に与えられた音声データに対する印象語の自動抽出方式を示した。本方式によって, 事前に印象語を付与した音声訓練データを用いたクラスタリングを行うことで新規で与えられた音声データについて, それぞれのクラスタ中心との距離計算を行うことで, その音声データに合致した印象語を抽出することが可能となる。

今後の課題としては, 利用ユーザの声の特徴量やクラスタごとの音声特徴量からユーザが声の印象をどのようにすれば変容させることができるのか, その方法や改善案を提案するシステム方式の実現, 複数人によるアンケート方式をとった上での, 音声訓練データの音声印象語付与などが挙げられる。

参考文献

- [1] 池本真知子, 鈴木直人, 感情表出時の声質評価尺度の作成, 感情心理学研究, 15(2), pp. 80-88, 2008.
- [2] 木戸博, 粕谷英樹, 通常発話の声質に関連した日常表現語の抽出, 日本音響学会誌, 55(6), pp. 405-411, 1999.