

## ウェブ収集発話を対象とした若年者判別の検討

宮森 翔子† 西村 竜一† 入野 俊夫† 河原 英紀† (和歌山大学 ‡)

## 1. はじめに

現在、子どもを有害な物や情報から守るため、様々な場面で年齢確認が行われている。また、顔画像などといった生態情報を用いた年齢確認の実用化に向けた検討が進められている。そこで、本研究では生態情報の一つとして、発話情報を用いた若年者自動判別を検討する。

なお、本研究は、若年者自動判別を目指す。大人と子どもの境界年齢を明確には定義していない。以下では、年齢閾値という概念を導入し、9歳から18歳まで1歳単位で年齢閾値を変化させ検討を進める。具体例を挙げると、年齢閾値が15歳の時は満年齢14歳までを子ども、15歳以上を大人とみなすことになる。

## 2. 若年者判別実験

本研究では、次の2種類の判別実験を行った。

- 人間の主観による若年者判別
- 機械による自動での若年者判別

2つの実験の結果を比較し、自動判別の目標値と、自動判別に用いる年齢閾値を検討する。

若年者自動判別システムの利用場所には一般家庭などが想定される。以下の実験では利用実態に即した調査を行うため、実環境のデータを用いるのが望ましい。したがって本研究では、一般家庭などにおいて録音された発話データの収集実験も行った。

## 2.1 大人・子ども発話のネットワーク収集実験

実環境データとなる利用者発話を、インターネット上に公開した音声収集実験ウェブサイトを通じて収集した。音声収集システムには、音声ウェブシステム w3voice<sup>1),2),\*</sup>を用いた。なお、発話者は楽天リサーチ社のモニタ誘引サービスを介して募集した。

図1に実験サイトの全体構成を示す。実験サイトは、練習1つ・本番2つの合計3回の録音ステップを持つ。また、録音ステップの終了後には、発話者の属性を調査するアンケートを用意した。

「練習」「本番1」「本番2」の録音ステップでは、発話者に簡単な質問が提示される。発話者は発話で回答する。我々は収録した発話データを実験サイトのサーバを介して収集した。各録音ステップの質問を以下に示す。

- 練習：「今日は晴れていますか？」
- 本番1：「好きな食べ物は何ですか？」
- 本番2：「好きな言葉を教えてください」

## 2.1.1 収集実験結果

実際に発話収集実験を行った。実施期間は、2009年2月25日から3月30日までである。ユニークIP数で2,011のアクセスを得た。そのうち、3つの録音ステップ

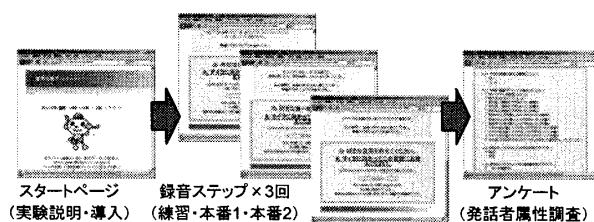


図1 実験サイトの全体構成

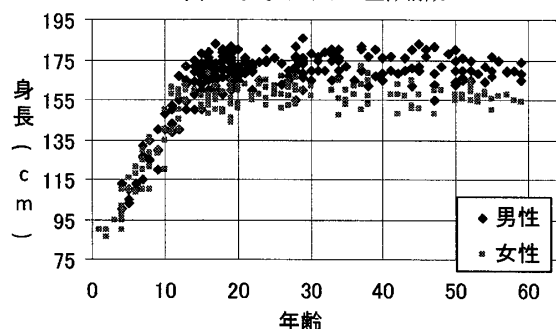


図2 発話者の年齢・身長分布

とアンケートを完遂できたものは432であった(回答率12.7%)。この中には無効な録音データ及びアンケートの入力ミスが含まれるため、作業員一人の手で内容を確認した。その結果、有効な発話数は1,109であった。発話者数は389(ユニークIP数384)であった。

図2に、アンケートの自己申告によって得られた年齢と身長の散点図を示す。横軸は年齢、縦軸は身長(cm)であり、各点は、赤は女性、青は男性の発話者を示す。

## 2.2 人間の主観による若年者判別実験 実験条件

収集発話の一部を用いて、人間の主観による若年者判別実験を行った。聴取者は5名(うち男性2名、女性3名)である。判別対象となる発話は、収集実験で収集された発話から抜粋した260発話(うち男声146、女声114)とした。判別対象の発話は、全て本番2(質問：好きな言葉を教えてください)に対する回答の発話である。また、発話者の年齢は2~59歳であった。聴取者は発話サンプルをスピーカーから聞き、各発話に対して次の3つの設問に答えることとした。

- 大人・子どものどちらに聞こえるか
- およそ何歳に聞こえるか
- 男女どちらの性別に聞こえるか

## 2.3 機械による若年者自動判別実験 実験条件

機械による若年者自動判別実験として、今回は話者認識に用いられる混合正規分布モデル(GMM)を音響モデルにした尤度比較を行った。

判別対象となる発話は発話収集実験で収集した1,109発話とした。まず収集発話を年齢及び性別に基づき、子ども、大人(女性)、大人(男性)の3クラスに分類した。学習の段階では、HTK 3.4.1を用い、各クラスに対して、音響特徴量を抽出し、GMM音響モデルを構築した。分析に用いた音響特徴量は、音声認識で用いるものと同様の12次元のMFCCと $\Delta$ MFCC、 $\Delta$ Powerである。構築

A proposal of children discrimination based on web collected utterances

† Shoko Miyamori, Ryuichi Nisimura, Toshio Irino and Hideki Kawahara

‡ Wakayama University

\*1 <http://w3voice.jp/>

した GMM の混合数は 128 である。判別段階では、入力発話に対し GMM 音響モデルの尤度を比較して最も高い尤度を得たクラスに分類する。判別に用いたデコーダは、Julius 4.1.2 である。

評価では、収集発話から評価用データを抜き出し、残りを学習用とする 10 分割の交差検定を行った。各発話者は 3 回の発話を行った<sup>\*2</sup>が、学習用データに評価用データの発話者を含めない条件 (話者オープン) である。

### 2.4 実験結果

人間の主観による判別と自動判別の各正解率を図 4 に示す。正解率は子ども発話を子どもとして判別した割合を示す。実線は自動判別による正解率、点線は人間の主観による判別での正解率である。横軸は年齢閾値である。

全体的に自動判別の正解率が低い。その中で、9 歳以下の子どもを検出する条件時 (年齢閾値 10 歳) にもっとも高い正解率 66.9% を得た。一方、人間の主観による判別では、年齢閾値 12 歳にもっとも高い正解率 87.7% となった。さらに、年齢閾値 13 歳では、人間の主観による判別と自動判別の正解率に最大の差 28.2% が見られた。

また、年齢閾値にともなう正解率の推移では、年齢閾値 15 歳以降において、両者とも正解率が下降している。この原因には変声期の影響が考えられる。この時期の声は音響的な特徴に変動が大きいため、人間の主観においても正確に判別するのが難しいことがわかった。

## 3. 子ども発話のネットワーク収集実験

収集実験で集まった発話データのうち、15 歳以下の子ども発話サンプル数は 291 発話であった。これは全体の 26% であり、大人・子どもの間にサンプル数の偏りが生じている。そのため、子ども発話を増やす必要があり、追加の収集実験を行った。以下では、追加の収集実験の結果を述べる。また、今後は、追加で収集した子ども発話を加えての自動判別実験を予定している。

### 3.1 子ども発話のネットワーク収集実験 実験条件

追加での子ども発話の収集実験では、発話者の対象を 0~19 歳とした。実験は前回と同様に、w3voice を用いた音声収集実験ウェブサイトを作成し、インターネット上に公開する形でを行った。サイトの構成や録音ステップ数は前回と同様である。なお、子どもが実験に飽きないようにするため、導入部にマンガや Flash アニメーションを用いた。各録音ステップの質問を以下に示す。

- 練習: 「こんにちは」と言ってください」
- 本番 1: 「好きな食べ物は何ですか?」
- 本番 2: 「好きな言葉を教えてください」

#### 3.1.2 収集実験結果

実施期間は、2009 年 12 月 8 日から 12 月 28 日までである。ユニーク IP 数で 3,767 のアクセスを得た。そのうち、3 つの録音ステップとアンケートを完遂できたものは 720 であった (回答率 19.1%)。また、内容を確認した結果、有効な発話数は 1,944 であった。発話者数は 661 (ユニーク IP 数 653) であった。図 5 に年齢ごとの発話者の人数を示す。グラフの赤色部分は女性、青色部分は男性の人

\*2 録音状態の不備により、実験に用いた発話の数 が 1 または 2 の発話者も存在する

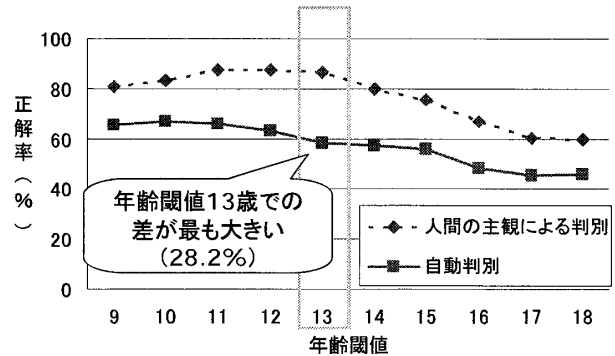


図 4 若年者判別実験 実験結果比較

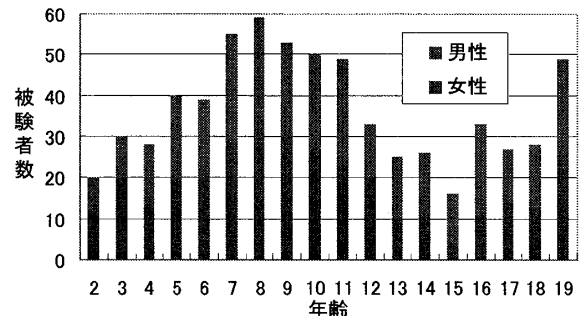


図 5 子ども発話収集実験 発話者数

数を示す。

この実験により、15 歳以下の子どものサンプル数は 1,533 発話となった。これはデータ全体 (3,053 発話) の 59.7% であり、発話サンプル数の偏りをなくすることができた。

## 4. まとめ

本研究では、発話による若年者自動判別を提案した。その実現を目指し、収集発話を使用した大人・子ども発話の人間による判別と自動判別の比較を行った。

判別結果の比較から、若年者自動判別の年齢閾値は 13 歳が妥当であると分かった。しかし、現段階において自動判別は、目標とする人間の主観による判別に対して精度が不足している。したがって、今後は自動判別に用いる特徴量として、言語的特徴も組み込んだ判別法<sup>3)4)</sup>の導入を検討する。

**謝辞** 本研究の一部は、科学研究費補助金及び和歌山大学オンリーワン創成プロジェクトの支援を受けた。

### 参考資料

- 1) 西村 他, “音声入力・認識機能を有する Web システム w3voice の開発と運用”, 情報処理学会研究報告, 2007-SLP-68-3, 2007.
- 2) Ryuichi Nisimura, et al., “Development of Speech Input Method for Interactive VoiceWeb Systems”, HCI International 2009, 2009.
- 3) Ryuichi Nisimura, et al., “Public Speech-Oriented Guidance System with Adult and Child Discrimination Capability”, Proc. ICASSP2004, Vol.1, pp.433-436, 2004.
- 4) 西村 他, “大人・子供に適応した音声情報案内のためのユーザ自動識別”, 電子情報通信学会技術研究報告, SP2003-129/NLC2003-66, 2003.