E-038

大人・子ども発話の自動識別に基づく安心 Web システムの検討

Proposal of safety web systems using adult and child voice discriminations

宮森 翔子† 西村 竜一† 鈴田 健太郎† 河原 英紀† 入野 俊夫†
Shoko Miyamori Ryuichi Nisimura Kentaro Suzuta Hideki Kawahara and Toshio Irino

1. はじめに

子どもを危険なウェブサイトから隔離するため，家庭等でのウェブフィルタリングの導入が進められている．本研究では，利用者の年齢層を PC のマイクで集音した発話に基づいて判別するウェブフィルタリングを提案する．

今回，提案システムの実現に向けて，(1) 音声ウェブシステム w3voice を用いた大人・子ども発話のネットワーク収集実験，(2) GMM 音響モデルを用いた若年者自動判別の予備実験を行ったので，その結果を報告する．

2. 大人･子ども発話のネットワーク収集

実環境利用者発話を収集するため，音声ウェブシステム w3voice[1),2),*1]による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，被験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，被験者に簡単な質問が提示される．被験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 収集実験結果

実際に発話収集実験を行った．なお，被験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．

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，作業者一名（大学生）の人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．被験者数は 389（ユニーク IP 数 384）であった．

2.2 被験者の年齢

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布図を示す．横軸は年齢，縦軸は身長（cm）であり，各点は，赤は女性，青は男性の被験者を示す．

本研究は，大人と子どもの自動判別を目指すが，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，8 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．具体的には，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を意味する．

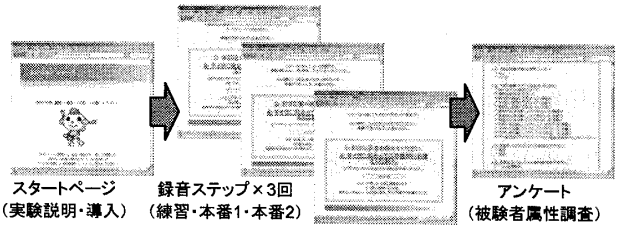

図 1 実験サイトの全体構成

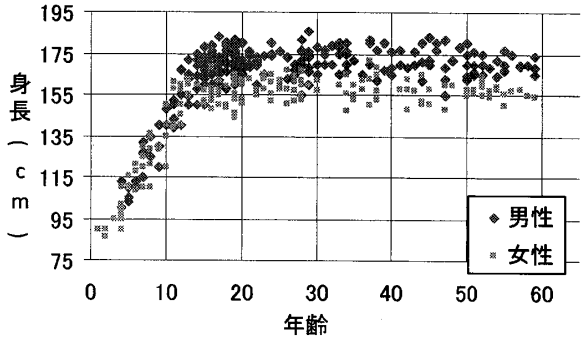

図 2 被験者の年齢・身長分布

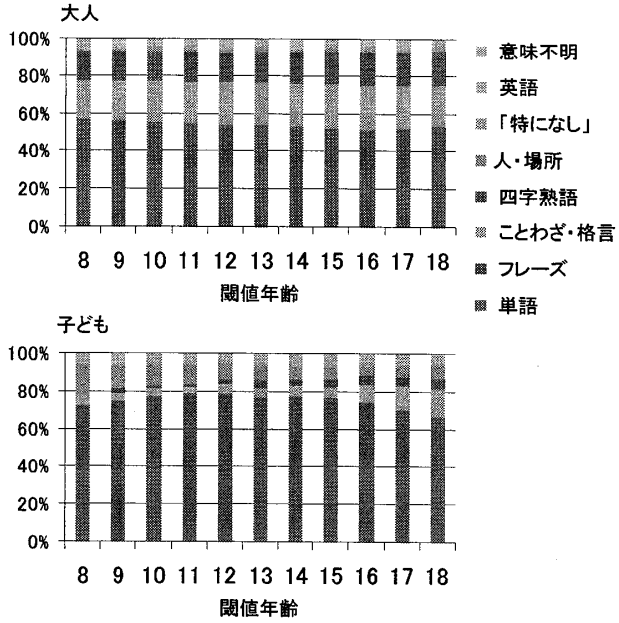

図 3 発話内容に関する分類結果

2.3 発話内容に関する分類

「本番 2（好きな言葉を教えてください）」に対し得られた回答発話を書き起こしし，その内容に応じて以下の 8 種類に分類した．作業は大学生一名が人手で行った．

† 和歌山大学システム工学部

*1 http://w3voice.jp/

- 単語：文章ではなく一般的な単語のみの回答.
- フレーズ：日常的な生活で使う文章やキャッチコピーのような文章の回答.
- ことわざ･格言：一般的に使われることわざや格言の回答. 漫画・書籍から引用した有名な文章も含む.
- 四字熟語：一般的な四字熟語を含む回答.
- 人・場所：特定の人物や地名等を含む回答.
- 「特になし」：好きな言葉が無いと答えた回答.
- 英語：英語による回答.
- 意味不明：作業員が意味のわからなかった回答.

図 3 に分類結果を示す. 大人と子どもの各グラフ上の割合は，収集発話を大人と子どもに分割した後の割合を示す. グラフの横軸は年齢閾値である.

発話内容に大人と子どもで異なる傾向があることがわかる. 子どもには「単語」の発話が，大人には「ことわざ・格言」や「四字熟語」の発話が多くみられた. これは，大人と子どもの判別基準に，発話の言語的特徴（単語や言い回し）を利用できることを示唆している.

3. GMM 音響モデルを用いた若年者自動判別

集めた発話を用いて簡単な大人・子どもの自動判別実験を行った. 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.

3.1 実験条件

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 ΔMFCC，ΔPower である. 構築した 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. 各被験者は 3 回の発話を行っている[*2]が，学習用データに評価用データの被験者を含まない条件（話者オープン）である.

3.2 実験結果

正解率を**図 4** に示す. 実線は子ども発話を子どもとして判別した正解率，点線は大人と子どもの全発話の正解率である. 横軸は年齢閾値である. 全体的に，子どもの正解率が低い. 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（このときの大人を含めた全体の正解率は 94.0%）.

年齢閾値 10 歳の詳細結果を**図 5** に示す. この結果から，大人に関しては，男性・女性ともに高い精度で判別できていることがわかる. 一方で，子どもに関しては，大人の女性と誤判断することが多い結果となった（23.1%）.

また，年齢閾値にともなう正解率の推移では，年齢閾値 11 歳以降において，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，判別が難しいと考えられる.

*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 の被験者も存在する.

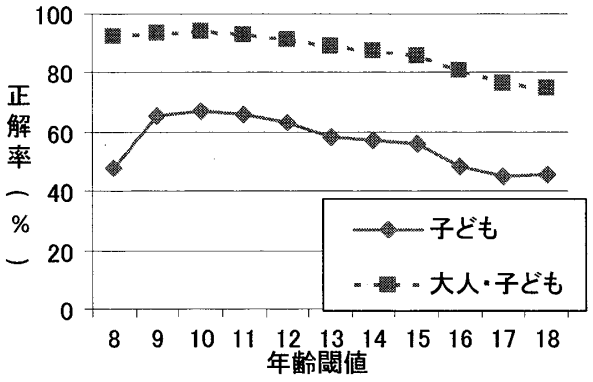

図 4 大人・子ども判別正解率

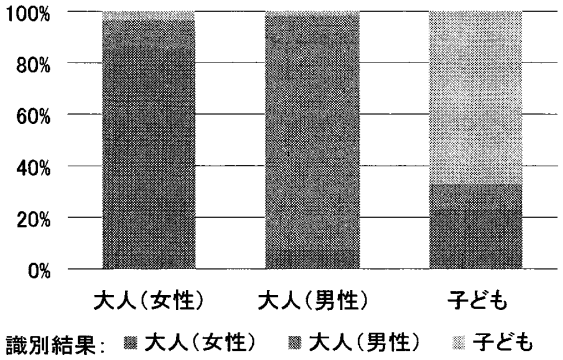

図 5 10 歳以下の子ども検出の結果詳細

4. まとめ

本研究では，発話による若年者自動判別を可能とするウェブフィルタリングサービスを提案した. その実現を目指し，（1）音声ウェブシステム w3voice を用いた大人・子ども発話のネットワーク収集実験，（2）GMM 音響モデルを用いた若年者自動判別の予備実験を行った. 発話に含まれる言語的特徴の分析と予備実験の結果から提案システムが実現可能であることを示した.

今後は，引き続き，発話の収集と詳細な分析を行う. また，若年者判別の精度が不足しており，精度向上が必要である. 今後は言語的特徴も組み込んだ判別法[3),4)]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. (発表予定)
3) Ryuichi Nisimura, et al., “Public Speech-Oriented Guidance System with Adult and Child Discrimination Capability”, Proc. ICASSP2004, Vol.I, 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.