様式 C-19

科学研究費補助金研究成果報告書

平成２３年　６月　１日現在

機関番号：１４７０１
研究種目：　若手研究（B）
研究期間：　2008 ～ 2010
課題番号：　20700090
研究課題名（和文）　若年話者を声で識別する安心ウェブシステムの研究開発

研究課題名（英文）　Development of Secure Web System with Voice Interface to Identify Child Speakers
研究代表者
西村　竜一　（NISIMURA RYUICHI）
和歌山大学・システム工学部・助教
研究者番号：00379611

研究成果の概要（和文）：本研究は，発話を入力とする大人・子ども識別法の開発を行った．(1) 音声ウェブシステムを用いた実環境発話収集 (2) HMM による識別法を備えたプロトタイプの作成 (3) 人間と機械の識別能力の比較 (4) 収集発話の言語特徴の調査を行った．しかし，実験から，変声期に当たる 10 代若者の音声を正しく判別できないことが明らかになった．この問題に対し，(5) HMM と SVM の組み合わせによる識別法を提案し，改善を得ることができた．

研究成果の概要（英文）：In this study, the web-based system identifying child speakers, which can be adopted in web filtering system to protect children from the dangers of the Internet, has been developed as follows: (1) The collection of actual utterances was organized by using our voice-enabled web system. (2) The prototype system having the HMM-based classifier was developed. (3) The HMM-based method and the human's hearing abilities in distinguishing between child and adult utterances were compared. (4) We considered the linguistic features that help distinguish between children and adults. (5) The classifier based on combining 24-class HMMs (Hidden Markov Models) and SVMs (Support Vector Machines) was proposed. In the experiments, we proved improvements in accuracies in identifying child speakers by using voices captured from real web users.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2008 年度	1,800,000	540,000	2,340,000
2009 年度	700,000	210,000	910,000
2010 年度	800,000	240,000	1,040,000
年度			
年度			
総　計	3,300,000	990,000	4,290,000

研究分野：音情報処理学
科研費の分科・細目：情報学・メディア情報学・データベース
キーワード：Ｗｅｂサービス，音声インタフェース，子ども，安全・安心，隠れマルコフモデル，サポートベクタマシン，フィールドテスト

１．研究開始当初の背景

ICT（情報通信技術）システムが社会基盤として普及するに従い，利用者の年齢を確認する技術に対するニーズは高まっている．あらゆる場面において子ども利用者の判別技術は求められてきた．例えば，タバコや酒，切符の自動販売機，他にも青少年向けの危険ウェブサイトのフィルタリング技術などが挙げられる．タバコや酒の購入ではもちろんのこと，青少年に悪影響を与える可能性のあるウェブページでは，子どもを保護する目的で年齢確認が必要である．また，これからの普

及が予期されるロボット等の対話型音声インタフェースにおいても，要素技術として年齢確認は有効である．具体的には，利用者の年齢層に応じてシステムの反応を切り替えることで，より柔軟で親切な対話処理を実現できると考えられている．

ユーザに過度な負担を与えることなく年齢確認を実現するには，生体情報を入力とする認識技術の応用が有効である．その一つの例として，2007 年には，顔認識で成人判別を行う機能を備えたタバコの自動販売機が話題となった．しかし，現実の実用化例では，子どもを大人と誤認識することを防ぐため，システムが大人とみなす年齢が意図的に高く設定されているなど，10 代から 20 歳前後を境界とした大人・子ども判別の技術は，まだ確立されていない．

２．研究の目的

本研究は，生体情報として発話を入力とする、大人・子ども自動識別サービスを ICT の技術基盤であるウェブ上に整備することが目的である。その実現に向け，特に，大人・子ども話者自動識別アルゴリズムについて検討を進めた．

生体情報として顔画像や体型，動作パターン等，様々な信号を利用することが考えられるが，本研究では，その中でも音声信号に着目した．自然な会話で生じる発話を入力とすることで，利用者に負担を与えることなく，会話を繰り返し，正確性の高い自動識別を実現できると考える．ここに生体情報として発話を用いることの利点がある．また，発話情報からは，音響的特徴と言語的特徴の 2 種類のパラメタを抽出することができる．これらを併用し，識別アルゴリズムの中に効率的に組み込むことができれば，自動識別の精度向上が期待できる．

３．研究の方法

本研究では，まず，発話を入力とする大人・子ども識別技術の実現可能性を検証するために，音声ウェブシステム w3voice を用いたプロトタイプの作成を行った．また，作成の過程では，大規模な実環境発話の収集・データベース整備及び人間と機械の識別能力の比較実験を行った．次に，提案システムの核となる大人・子ども自動識別アルゴリズムの改良を進めた．プロトタイプに用いたベースラインのアルゴリズムは，既存の音声認識を転用し，実装した．しかし，音響的特徴のみを扱った手法では，特に，変声期に当たる 10 代の若者の音声を，子どもとして正しく判別することはできないことが実験によって明らかになった．変声期の音声は，人間の耳でも判断することが難しい．この問題に対処すべく，本研究では，HMM(隠れマルコフモデル）と SVM(サポートベクタマシン）を 2 層に組み合わせた判別手法を検討した．下記に，各実施項目をまとめる．

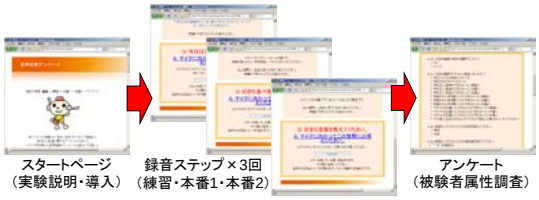

図 1 発話収集に用いたウェブサイトの構成

(1) 音声ウェブシステムによる実環境発話の収集

大人・子ども判別の技術を応用する際，家庭の利用がまず想定されるため，本研究では，家庭において PC に向かって発声された音声を対象として議論を進めることが望ましい．本当の家庭環境で PC に向かって発声した音声を集めた事例はこれまで限られており，特に，子どもによる発話を含んだデータベースが整備されたことは皆無である．そこで，我々の音声ウェブシステム w3voice を用いて，インターネットを介した発話の収集を試みた．音声ウェブシステム w3voice は，利用者によって録音された音声をサーバに自動送信する機構を持つため，サーバ上で音声を収集することができる．

音声収集用に我々が構築したウェブサイト（図 1）では，利用者が発話を行う過程として，「練習」「本番 1」「本番 2」の 3 ステップが用意されており，各ステップには簡単な設問が用意されている．本番 1，本番 2 の設問内容を以下に示す．

- 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.

発話者は各ステップにおいて設問への解答を発話し，録音する．全ての録音ステップが完了した後，発話者は，自身の属性及び使用した機材に関するアンケートに回答する．なお，低年齢の発話者には保護者が付き添い操作を行うように保護者に依頼した．発話者が低年齢の際は，録音時の PC 操作及びアンケートの回答は，付添いの保護者が代行するように事前に要請した．

(2) プロトタイプシステムの作成

本研究では，初期段階で，発話の音響的特徴に基づいて大人・子ども自動判別を行うプロトタイプを作成した．研究を推進する上で，目で見て動作が把握できるように，検証対象となるシステムが存在する方が議論を円滑に進めることができるためである．

(3) 収集発話と用いた人間と機械の大人・子ども識別能力の比較

本研究を進めるにあたって，「そもそも人間

の耳は発話から大人と子どもを聞き分けることができるのだろうか」という疑問が生じる．大人・子ども自動判別技術の検証を進める上で，その精度の目標値と，大人と子どもの境界となる年齢閾値を検討するため，人間による主観と自動識別の能力を比較した．

人間の主観による判別実験に協力した被験者は，男性 2 名，女性 3 名である．対象とした発話は，収集発話のうち本番 2（質問「好きな言葉を教えてください」）で録音された 260 発話（男 146，女 114 ）である．ここでの比較の対象としては，プロトタイプシステムでも採用している HMM による自動識別を用いた．HMM の学習には，2,361 発話を用いた．

(4) 収集発話に対する言語特徴（使用単語傾向）の調査

本研究では，録音信号の音響的な側面に加え，言語的な特徴を識別アルゴリズムに組み込むことができることに，発話を入力とする利点があると考える．収集発話に含まれる言語的な特徴の一つとして，大人と子どもで使用する単語の傾向を調査した．発話収集のうち，本番 2 で被験者に提示した質問「好きな言葉を教えてください.」に対して得られた回答の書き起こしを，その内容に応じて 8 分類した．作業は一人の人間（大学生）が人手で行った．分類に用いたのは，「単語」「フレーズ」「ことわざ・格言」「四字熟語」「人・場所」「特になし」「英語」「意味不明」の 8 種類である．

(5) HMM+SVM の 2 層による大人・子ども識別アルゴリズムの検討

実験の結果，前述のプロトタイプシステムの手法では，変声期に当たる 10 代若年者の識別精度が低く，本研究の目的から考えて実用的では無かった．少数のクラス間の単純な尤度比較のみで識別結果を得ることに原因があると考えた．そこで，HMM に加え，SVM を用いる 2 層の識別アルゴリズムを検討した．提案手法は，HMM のクラスを 24 に増やすことで，HMM が出力する尤度を 24 個に増やすことができる．取得した尤度を単純に大小比較するのではなく，24 次元の入力特徴とみなすことで SVM による二値識別を適用し，発話者を大人・子どもに判別することにした．

本手法は，1 層目で HMM の構築と尤度の算出を行い，2 層目で用いる特徴を構成する．発話者の年齢と性別に基づき，収集発話を男女 5 歳ごとの 24 クラスに分類した．発話データに周波数分析等を適用し，音響特徴量を抽出した．音響特徴量から 3 状態の HMM（24 クラス）を構築した．すべての発話を用いて 24 クラス HMM に対する尤度を算出した．得られた尤度を発話のフレーム数で正規化した．正規化音響尤度と HMM のクラス番号をセットにし，SVM の入力特徴を構成した．

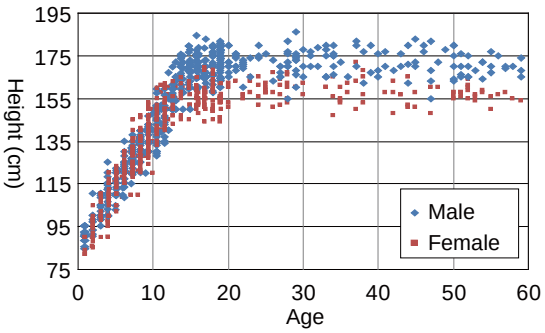

図 2 発話者の身長・年齢分布

2 層目では，1 層目で作成した特徴を入力にした SVM による 2 クラス識別を行う．1 層目で得た尤度を組み合わせて SVM の入力データを作成した．入力データに対し，発話者の年齢と年齢閾値に基づいた 2 クラスラベル（年齢閾値以上(大人）をポジティブ，年齢閾値未満(子ども）をネガティブ）を付与した．作成した入力データとラベルに基づき SVM による 2 クラス識別を適用した．

４．研究成果

(1) 音声ウェブシステムを用いて実環境発話を収集した結果，用意した収集用ウェブサイトに，ユニーク IP アドレスで 5,778 のアクセスを得た．そのうち，3 つの録音ステップを完遂した発話者は 1,152 名であり，回答率は 19.9%であった．収集された発話の中には無効な録音データやアンケートの入力ミスなどが含まれるため，大学生 2 名が人手で内容を確認した．その結果，発話者 1,050 名分の 3,053 発話が有効であった(1,037 ユニーク IP アドレス)．

図 2 に，収集した発話の発話者の年齢と身長の散布図を示す．図の赤点は女性の発話者，青点は男性の発話者を示す．全ての発話サンプルのうち，15 歳以下の子どもの発話サンプルは 1,533 発話であり，全体の 59.7%を占めた．ただし，10 歳未満の発話者に対して，10 代の発話数は少ない．特に，15 歳によって発話されたサンプルは 26 発話であった．よって，10 代の発話を追加収集する予定ではあるが，今のところ，収集発話における発話者の年齢による偏りは存在する．

(2) ウェブ上で動作するプロトタイプシステムは，通常のウェブシステムに音声入力インタフェースを付加するフレームワークである w3voice システムを用いて実装した．これにより，本システムは，クラウド型のウェブシステムとして動作する．ブラウザ上で録音された発話は，我々の研究室に設置されたサーバに自動的に送られる．次に，サーバ上のプログラムが自動識別をし，結果をブラウ

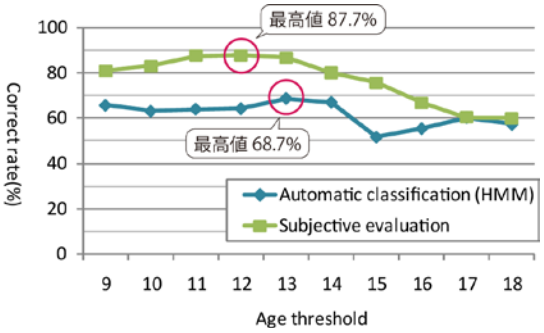

図 3 人間と自動識別の正解率比較

ザに出力する．

識別処理には，既存の音声認識プログラムを転用した．統計モデルの学習では，まず，収集発話を発話者の年齢と性別に基づき，「大人男性」，「大人女性」，「子ども」の 3 クラスに分けた．次に，各クラスの発話から音響特徴量を抽出し，3 状態の HMM を学習した．音響特徴量には，12 次元の MFCC，ΔMFCC，ΔPower を用いた．識別の段階では，HMM の尤度比較により，最も高い尤度を得たクラスを結果とする．この処理には音声認識プログラム Julius を用いた．

(3) 図 3 に，子ども発話に対する，人間の主観とプロトタイプで用いた自動識別法の正解率の比較を示す．正解率は，年齢閾値に満たない年齢の話者音声が，子どもであると正しく判別された割合を示す．図の緑線は人間の正解率，青線は自動識別の結果を示す．横軸は年齢閾値である．なお，本研究では，大人と子どもの境界となる年齢を示す値として年齢閾値を定義して用いた．例えば，年齢閾値 15 歳の場合は，15 歳未満の発話者を子ども，15 歳以上の発話者を大人とみなすことになる．実験では，年齢閾値を 1 歳ごとに変化させて，年齢閾値の変化によって生じる識別能力の違いを議論した．

全体的に，自動識別の結果は，人間よりも正解率が低い．また，ここで問題になるのは，どちらの場合においても，年齢閾値 15 歳以上では正解率が低下していることである．この傾向の一因には，変声期の影響が考えられる．変声期における音声には，音響的に大きな変動がある．そのため，人間にとっても年齢層の判別が難しいことが予想される．しかしながら，提案システムの実用化を想定すると，年齢閾値 10 代後半に要求が存在する．このため，15 歳以上での精度低下は，解決を要する重要な課題であると確認した．

(4) 図 4 に収集発話に対する言語特徴（使用単語傾向）の調査結果を示す．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

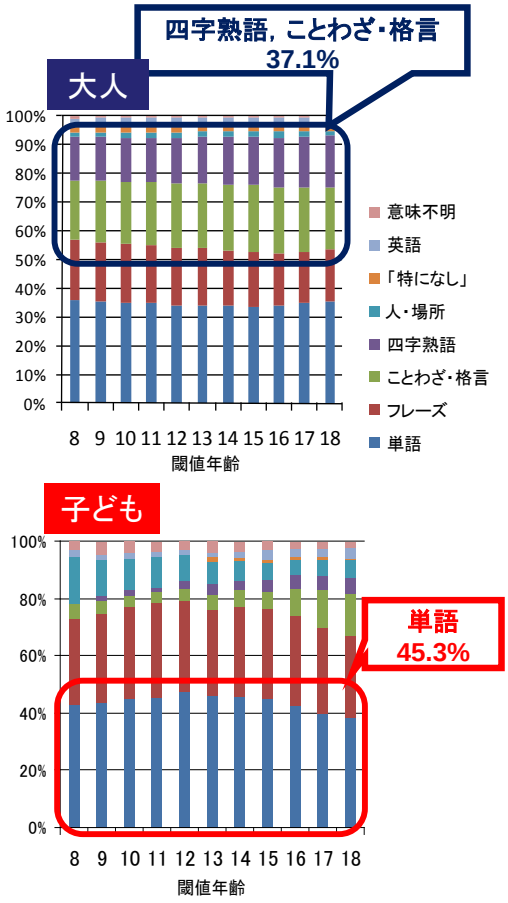

図 4 単語出現傾向の比較

軸は年齢閾値である．この結果より，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．子どもは「単語」の発話が多い．一方，大人は「ことわざ・格言」や「四字熟語」が多くなる．具体的には，年齢閾値 15 歳の時の大人発話における「ことわざ・格言」の割合は 23.6%に対し，子どもは 6.2%であった．「単語」に関しては大人 33.4%，45.4%となった．この結果は，大人と子どもを自動識別する際の特徴量に，発話の言語的特徴（単語や言い回し）を利用することが有効であることを示唆している．

(5) 本研究の提案法となる HMM+SVM の 2 層による大人・子ども識別法を評価した．図 5 に正解率を示す．子ども判別の正解率は，11 歳以上の年齢閾値において，常に 60%以上を示した．大人を判別した場合においても，年齢閾値 17 歳まで 60%以上の正解率を示す結果が得られた．また，F 値においては，年齢閾値 11 歳から 17 歳までの区間では，0.7 以上を示す結果となった．また，年齢閾値 13 歳のとき，正解率 78.9%を得ることができた．これは，プロトタイプシステムで用いた HMM による識別法（従来法）と比較して，10.2 ポイントの精度向上である．

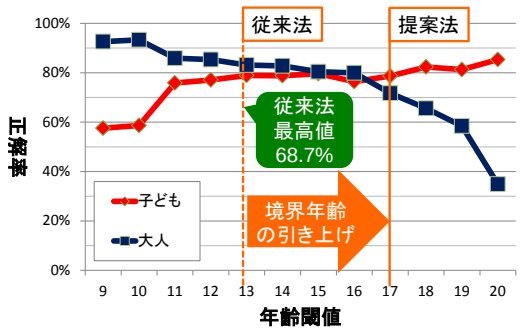

図 5 評価実験結果（提案法の正解率）

また，HMM では，年齢閾値 13 歳のときに最高値だったのに対し，提案法は 17 歳でも高精度を得た．つまり，提案法は，前述の実験では検討を要する課題として残った，年齢閾値の向上に対し，有効な改善を得たことがわかる．従来は，変声期の 10 代を含む高い年齢閾値における精度に問題があった．提案法により，大人の判別性能を維持したまま，10 代以降の子ども判別性能を上げるのに成功した．以上にように，本研究は，実用化に向けた着実な成果を得ることができた．

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔学会発表〕（計 19 件）

① 宮森翔子，西村竜一，栗原理沙，河原英紀，入野俊夫，“実環境発話を用いた子ども判別アルゴリズムの検討”，日本音響学会 2011 年春季研究発表会，2011 年 3 月 10 日，早稲田大学（東京都），査読無.

② Shoko Miyamori, Ryuichi Nisimura, Lisa Kurihara, Toshio Irino, Hideki Kawahara, “Comparing Abilities of Humans and Machine for Child Speaker Identification based on Web Utterances Collection”, 2nd APSIPA Annual Summit and Conference (APSIPA2010) Student Symposium, 2010 年 12 月 14 日, Biopolis（シンガポール），査読有.

③ Shoko Miyamori, Ryuichi Nisimura, Lisa Kurihara, Toshio Irino, Hideki Kawahara, “Real world utterance collection using voice-enabled web system for child speaker identification”, 13th Oriental COCOSDA, 2010 年 11 月 25 日，カトマンズ（ネパール），査読有.

④ 宮森翔子，西村竜一，栗原理沙，入野俊夫，河原英紀，“ちょっとした一言の音声認識による子ども利用者判別法の検討”，FIT2010 第 9 回情報科学技術フォーラム，2010 年 9 月 7 日，九州大学（福岡県），査読無.

⑤ 栗原理沙，西村竜一，宮森翔子，入野俊夫，河原英紀，“音声ウェブシステムを用いて収集した実環境子供発話に関する調査”，FIT2010 第 9 回情報科学技術フォーラム，2010 年 9 月 7 日，九州大学(福岡県），査読無.

⑥ 宮森翔子，西村竜一，入野俊夫，河原英紀，“ウェブ収集発話を対象とした若年者判別の検討”，情報処理学会創立 50 周年記念（第 72 回)全国大会，2010 年 3 月 11 日，東京大学（東京都），査読無.

⑦ Kentaro Suzuta, Ryuichi Nisimura, Hideki Kawahara, Toshio Irino, “Topic-Dependent Language Modeling for VoiceWeb Systems”, WESPAC X 2009 (The 10th Western Pacific Acoustics Conference), 2009 年 9 月 23 日，フレンドシップホテル（中国・北京），査読有.

⑧ Ryuichi Nisimura, Jumpei Miyake, Hideki Kawahara, Toshio Irino, “Development of Speech Input Method for Interactive VoiceWeb Systems”, Lecture Notes in Computer Science (HCII2009), vol.5611, pp.710-719, 2009 年 7 月 22 日，サンディエゴ(アメリカ)，査読有.

⑨ 西村竜一，宮森翔子，鈴田健太郎，河原英紀，入野俊夫，“安心ウェブの実現に向けた大人・子ども発話のネット収集実験”，情報処理学会音声言語情報処理研究会，2009 年 7 月 18 日，飯坂温泉（福島県），査読無.

⑩ Ryuichi Nisimura, Jumpei Miyake, Hideki Kawahara, Toshio Irino, “Speech-to-text input method for Web system using Javascript”, IEEE SLT, pp.209-212, 2008 年 12 月 17 日，ゴア(インド），査読有.

⑪ 西村竜一，鈴田健太郎，河原英紀，入野俊夫，“音声認識 Web システムにおける単語辞書構築技術”，日本音響学会 2008 年秋季研究発表会，2008 年 9 月 12 日，九州大学（福岡県），査読無.

〔その他〕
http://w3voice.jp/

６．研究組織
(1)研究代表者
西村　竜一（NISIMURA RYUICHI）
和歌山大学・システム工学部・助教
研究者番号：00379611

(2)研究分担者
（　　）

研究者番号：

(3)連携研究者

(　　)

研究者番号：