社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2007－ＨＩ－122 (7)
2007－SLP－ 65 (7)
2007／2／9

PodCastleの提案: 音声認識研究2.0を目指して

後藤 真孝　　緒方 淳　　江渡 浩一郎

産業技術総合研究所
m.goto [at] aist.go.jp

あらまし　本稿では，Web 2.0に基づくWebサービスを提供し，音声認識性能の現状を積極的に開示することで，不特定多数のユーザの協力を得て音声認識技術を発展させていく研究アプローチ「音声認識研究2.0」を提案する．我々は，これを具現化した音声認識のキラーアプリケーションを目指して，音声認識に基づくポッドキャスト検索サービスPodCastleの公開を開始した．PodCastleでは，ユーザがWeb上の日本語のポッドキャストを全文検索できるだけでなく，認識結果の全文テキストも閲覧でき，さらに誤認識箇所を容易に訂正することもできる．これにより，ユーザが利用しながら訂正すると認識性能と検索性能が向上し，さらなる利用が促せるというポジティブスパイラルが生じることが期待できる．

PodCastle: A Proposal for Speech Recognition Research 2.0

Masataka Goto　　Jun Ogata　　Kouichirou Eto

National Institute of Advanced Industrial Science and Technology (AIST)

Abstract　In this paper, we propose *"Speech Recognition Research 2.0"* a research approach that provides users with a web service based on Web 2.0 to disclose state-of-the-art speech recognition performances and to promote speech recognition technologies in cooperation with anonymous users. In the quest for a killer application of speech recognition which embodies this approach, we launched a public web service *"PodCastle"* for searching podcasts on the basis of speech recognition. PodCastle enables users to accomplish a full-text search of Japanese podcasts on the web, read full texts of their speech recognition results, and easily correct recognition errors. We can thus expect a positive spiral where the improvement of the recognition and search performances through the correction by users encourages further usage of the web service.

1　はじめに

エンドユーザ（音声認識利用者）は，音声認識が有用な技術であることを実感していない．音声認識研究者は，音声認識が高度な技術に基づいており，想定した音声入力に対しては高い性能を示すことを知っている．そして，どのような音声が認識しやすいかも理解している．実際に，適切な条件下では，現在の音声認識の性能は一昔前と比べて驚くほど高い．一方，エンドユーザは音声認識の原理を知らず，過去に経験した範囲内で音声認識の有効性を判断している．そのため，音声認識にとってどのような音声が簡単で，どのような音声が難しいかは，充分には理解していない．そして，過去に自分の音声が正しく認識されなかった経験等があると，そのときの印象で，音声認識の有効性に疑問を抱き，使わなくなることが多い．様々な研究により音声認識率が向上し，文献1) で「なぜ音声認識は使われないか」が分析されたときの状況から進展してはいるものの[2)]，近年も文献3)~5) のような議論があったように，多くのユーザが利用に至らないという問題は依然として解決されていない．

本研究の第一の目的は，この問題を解決すべく，エンドユーザに現在の音声認識の技術レベルを把握してもらい，音声認識の普及と実用化を促進することにある．そこで我々は，不特定多数のユーザがポッドキャストを検索・閲覧できるWebサービス「PodCastle」（ポッドキャッスル）を公開し，様々な音声の認識結果の全文テキストをユーザと共有することを可能にする．ポッドキャストは，音声版のブログ (Weblog) に位置付けられ，Web上の音声データとして多数公開されている．そのため，ユーザ自身が発声しなくても，様々な音声認識結果を閲覧することで，認識技術の現状が把握できる．例えば，マイク入力した自分の音声が誤認識されると，それを不快あるいは恥ずかしいと思い，発声したがらないユーザがいるが，既に公開されているポッドキャストの認識結果を見てもそうした問題がなく，利用に躊躇がない．

しかし，ポッドキャストの内容や収録環境は多種多様であり，現在の音声認識技術ではそのすべてを適切に認識することはできない．こうした問題に対する音声認識分野での典型的アプローチは，研究者が認識対象の音声データを大量に収録してコーパスを作成し，書き起こしテキストを用意して学習・適応する方法である．ただし，このアプローチでポッドキャス

(i) ユーザが音声認識を体験することで，その性能を理解する．
(ii) 音声認識の性能向上にユーザが貢献する．
(iii) 性能が向上したら，それがより良いユーザ体験に結びつく．

図 1: 「利用される音声認識」へ向けたポジティブスパイラル（(i)～(iii)の各段階が繰り返される好循環）

トの全文検索を実現しようとすると，事実上，あらゆる音声に対するコーパスを整備する状況に近くなり，コストや時間の観点からも現実的でない．

本研究の第二の目的は，この問題を解決すべく，事前に対象となるコーパスを用意する考えを捨て，不特定多数のユーザの力を借りて音声情報検索と音声認識の性能向上を実現することにある．現在の音声認識技術でポッドキャストのような実世界の音声データを認識すれば，当然多くの誤認識箇所が発生する．そこでPodCastleでは，ユーザにそうした誤認識を訂正する協力をしてもらい，適切に検索できるようにしていく．さらに，その訂正履歴を学習に利用することで，運用中に自動的に音声認識の性能向上が図れる仕組みを実現する．これは，ユーザに「音声認識を育ててもらう」アプローチと言える．

本稿では，このように「ユーザに対して音声認識の現状を積極的に開示し，ユーザの協力を得て音声認識技術を発展させていく研究アプローチ」を「音声認識研究2.0」と名付ける．これにより，研究分野全体での問題意識の共有を図り，問題解決へ向けて力を合わせて取り組んでいけることを狙っている．これは，Web 2.0[6]を意識して付けた名称であり，Web 2.0の特長を取り入れた研究アプローチとも言える．以下，2章で音声認識研究2.0で提案する研究アプローチについて議論し，3章でその実例としてWebサービスPodCastleを提案する．そして，4章でWeb 2.0との関連を考察し，今後の展望を議論する．最後に，5章で本研究の意義を総括する．

2 音声認識研究2.0

「音声認識研究2.0」とは，不特定多数のエンドユーザの協力を仰ぎながら，音声認識の性能向上と実用化（利用率の向上）を共に実現することを目指した，音声認識の新たな研究アプローチである．そして，その実現のために，図1のポジティブスパイラルを回すことを提案する．従来は，音声認識の普及のために重要なこの三段階のそれぞれに阻害要因があったため，これが回っていなかったと考えられる．そこで，以下では各段階の抱えていた問題を順に考察する．

- (i)の性能理解に関しては，従来は，ユーザ自身の発声を認識した結果を見て，性能を誤解する可能性が高かった．多くの音声認識研究者は，ユーザとしての自分の発声ではなく，他人の適切な発声（コーパス中の音声）を認識した結果を目にする機会が多いため，実体験としても性能を誤解することはなかった．しかし，ユーザは何度か自分の発声が認識されない体験をするだけで，他の人の音声も同様に認識されないものだと誤解することがあった．
- (ii)の性能向上に関しては，従来，話者適応のためにユーザに例文を発声させたり，未知語を辞書登録させたりすることが多かった☆1．しかし，そうしたエンドユーザによる性能改善が，他のユーザと共有されて再利用されることはなく，総体としての音声認識の性能向上には，音声認識研究者しか貢献できなかった．そのために，不特定多数のユーザが共に性能向上を実感して，それに共同で貢献していくことを動機付ける要因はなかった．
- (iii)のユーザ体験向上に関しては，音声認識研究者の手元で日々性能が向上していても，その高い性能をユーザが体験する機会は非常に限られていた．例えば，研究目的で音声認識ソフトウェア（例えば文献7)）が公開されても，主に開発者向けでエンドユーザが直接利用する機会は少なく，音声対話システム（例えば文献8)）が街中に設置されても，その地域を訪れたユーザしか体験できなかった．音声認識を利用した市販ソフトウェアでも，数ヶ月～数年のバージョンアップのサイクルでしかユーザは性能向上を体験できなかった．

音声認識研究2.0では，これらの各問題を解決することで，図1のポジティブスパイラルを回し，音声認識を取り巻く状況を変革することを目指す．従来の典型的な研究アプローチ（以下，「音声認識研究1.0」と呼ぶ）との対比を表1に示す．ここでは対比する関係上，従来の研究アプローチを「音声認識研究1.0」と名付けているが，それは決して劣るものでも不要なものでもなく，今後の音声認識の発展のために継続して研究することが必要不可欠であることは間違いない☆2．これは，「音声認識研究1.0」を土台として，それに加えて「音声認識研究2.0」のアプローチにも取り組むべきであるという提案である．なお，音声認識の手法自体について議論しているのではなく，研究の方法論，アプローチについて議論しているため，「音声認識2.0」ではなく「音声認識研究2.0」と名付けた．

☆1 ただし，ユーザに意識させることなく利用中に自動的に話者適応したり，研究レベルでは未知語を自動獲得したりできるシステムも存在する．しかし，いずれの場合も，それらがエンドユーザ間で共有されることはなかった．

☆2 もちろん我々自身も，音声認識研究2.0によって難易度の高い音声データに対する性能上の問題点をより一層自覚することで，音声認識研究1.0に継続的に取り組んでいく．

以下，表1の項目について説明しながら，図1が実現できることを述べる．

- 音声認識研究2.0では，コーパスに基づいて学習した音声認識システムをディクテーションや音声対話等のスタンドアロンアプリケーションとして提供するのではなく，ポッドキャスト等の**Web上の音声データ**を対象に，ユーザが直接**検索・閲覧**できる**Webサービス**を実現する．これにより，図1(i)の性能理解が促進される．
- しかし，Web上の音声データを対象とすると，話題が従来の音声認識研究のようには限定できず，コーパスやその書き起こしも整備されていないため，多くの誤認識が起きる．また，認識用の辞書に登録されていない未知語も多くなる．それに対して音声認識研究2.0では，**話題非限定**な状況で多様な音声データの認識に挑戦し，誤認識箇所はユーザに訂正してもらって検索可能にする方針をとる．つまり，各音声データの検索用**アノテーション**として，書き起こしに相当する全文テキストをユーザの協力により整備していく．ここで重要なのは，その訂正内容を学習することで，まだ訂正していない部分や他の音声データに対する認識結果が改善される点である☆3．未知語に関しても，ユーザがまだアノテーション（訂正）していない**未アノテーション語**に過ぎないと考え，ユーザの訂正後に学習して語彙を増やしていく．このように，専門家である研究者だけでなく，ユーザ自身も訂正作業により図1(ii)の性能向上へ貢献することができる．
- さらに，これを個人的な訂正作業に留めずに，この**ユーザ参加**型の仕組みを発展させ，不特定多数のユーザの訂正結果をWebサービス上で共有して性能改善を図る**社会的訂正**の枠組みも提案する．社会的訂正では他の人々の利便性に貢献している実感が得られ，仮に一人では積極的に訂正する気持ちにならないとしても，他の人々が訂正している活動を見ることで，訂正して貢献する気持ちになる可能性がある．これは**集合知**(wisdom of crowds)を利用して図1(iii)のユーザ体験向上を実現するものである．

つまり音声認識研究2.0は，いわば**永久にベータ版**(perpetual beta)とも言える完全ではない音声認識に基づくWebサービスを，Web上で多数のユーザの協力を仰ぎながら使ってもらうことで機能改善し，研究を進めていくアプローチとして位置付けられる．

我々は，上記の音声認識研究2.0を具現化すべく，2006年1月にPodCastleプロジェクトを開始した．本プロジェクトでは，音声認識研究2.0とWeb 2.0の両者の考え方に基づくWebサービスであるPodCastleを中心に，図1のポジティブスパイラルを回していくことを目指している．

☆3 この点が，Web 2.0にはない，音声認識研究2.0の大きな特長である．例えば，Wikipedia[9]等の集合知を利用した他のWebサービスでは，ユーザの貢献は編集した項目に限定され，自動的に他の項目へ波及して改善されることはない．

表1: 従来の音声認識研究のアプローチ「音声認識研究1.0」と本研究で提案するアプローチ「音声認識研究2.0」の対比

音声認識研究1.0	音声認識研究2.0
スタンドアロンアプリ	Webサービス
ディクテーション	検索・閲覧
コーパス	Web上のデータ
話題限定	話題非限定
書き起こし	アノテーション
未知語	未アノテーション語
専門家参加	ユーザ参加
個人的訂正	社会的訂正
個人知	集合知
完成版	永久にベータ版

上記は，Web 1.0とWeb 2.0を対比した文献6)の表に影響を受けて記述した．これらの項目を満たすほど音声認識研究2.0的な研究事例と言えるが，Web 2.0の場合と同様に，すべてを満たさなければならないわけではない．

3 音声認識に基づくポッドキャスト検索サービス PodCastle

PodCastleは，ポッドキャストをテキストで検索，閲覧，編集できるソーシャルアノテーションシステム[10]であり，同時にWebサービスの名称でもある．近年，計算機上の音楽プレーヤやiPod等の携帯型音楽プレーヤで，Web上からダウンロードした音声データを効率良く聞く仕組み「ポッドキャスティング」が普及しつつある．そこで配信されるポッドキャストでは，一連のエピソードと呼ばれる音声データ（MP3ファイル）に加え，その流通を促すために，ブログなどで更新情報を通知するために用いられているメタデータRSS (Really Simple Syndication) が必ず付与されている．エピソードは作成者（ポッドキャスタ）側で任意のタイミング（毎日，毎週等）で追加できる．この仕組みによりポッドキャストは音声版ブログとも言われ，個人による音声データの発信，流通，入手が容易にできる点が普及を促してきた．そして，Web上のテキストに対して全文検索サービスが不可欠になったのと同様に，音声データに対してもPodCastleのような全文検索サービスの重要性が増している．

ポッドキャストを音声認識によりテキスト化し，ユーザがWebブラウザ上で入力した検索語を含むポッドキャストの一覧を提示できるWebサービスとしては，英語を対象に，既にPodscope[11]とPodZinger[12]が2005年から公開されている．Podscopeでは，ポッドキャストのタイトルだけが列挙され，音声認識結果のテキストは一切表示されないものの，検索語が

出現する箇所は再生できる．一方，PodZingerでは，これに加え，検索語が出現した周辺のテキストも表示され，ユーザが内容を把握しやすくなっている．それに対して，PodCastleは初めて日本語のポッドキャストに対する全文検索を実現するものであるが，仮にこれらが日本語に対応したとしても，以下の三つの点で本研究との違いは大きい．

1. 従来は，せっかく音声認識をしていても，表示されるテキストは一部に限定されており，音声を聞かずにポッドキャストの詳細な内容を視覚的に把握することはできなかった．
2. 音声認識により索引付けされた全文テキストは内部に隠蔽され，テキストに基づく他の全文検索Webサービスからは検索できなかった．
3. 音声認識にとって不可避な認識誤りが起きて検索に悪影響を与えていても，ユーザがそれらを訂正して改善することは不可能だった．

このように音声認識結果の完全開示による外部の検索サービスからの利用や，不特定多数のユーザの協力に基づく音声認識性能の向上を可能にするのは，本研究が初めてである．

3.1 PodCastleの3つの機能

PodCastleでは，「検索」，「閲覧」，「編集」の3つの機能を提供するWebサービスを一般公開しながら研究を進めることで，表1の音声認識研究2.0のすべての項目を満たし，図1のポジティブスパイラルを回していく．図1の(i)の性能理解は，「検索」機能と「閲覧」機能によって実現され，(ii)のユーザによる性能向上への貢献は，「編集」機能によって実現される．(iii)のより良いユーザ体験に結び付けるための性能向上については，訂正結果に基づく音響モデル，言語モデルの再学習や，未知語の自動辞書登録等の様々な手法に取り組んでおり，紙面の制約から，詳細は文献10),13)に譲って省略する．以下，これら3つの機能の特長を述べる．

3.1.1 「検索」機能

音声認識結果（およびユーザがそれを編集した結果）の全文テキストを索引情報として使用して，全文検索する機能である．一般的なテキスト全文検索サービスのように検索語をタイプすると，その語を含むエピソードの一覧が検索語付近のテキストと共に表示され，個々を試聴できる．そのうち一つを選択すると，次の「閲覧」機能に移行して全文テキストを読むことができる．

3.1.2 「閲覧」機能

ユーザが，検索したポッドキャストを「聞く」だけでなく，テキストで「読む」ことができる機能である．これにより，音声再生環境がなくても内容が把握でき，内容に関心があるかどうかを音を聞かずに判断できる．表示では，音声認識時に推定した形態素ごとの信頼度に応じて着色し，誤りを発見しやすくする．さらに，音声の再生に同期してテキスト中のカーソル（ハイライト）も動く．

図 2: 「音声訂正」[14] に基づく音声認識誤り訂正用インタフェース（通常の認識結果の下に競合候補が提示される）

このように各エピソードの全文テキストは外部公開されているため，外部の検索サービスで全文検索する際に，通常のWebページと共にPodCastleのエピソード閲覧ページが発見される．その結果，ポッドキャストがより多くのユーザの目に触れて価値が高まる．これはポッドキャスタにとってもメリットがあるので，不特定多数のユーザに加え，ポッドキャスタ自身も次の「編集」機能で訂正する動機付けの一つとなる．

3.1.3 「編集」機能

ユーザが検索・閲覧中に認識誤りを発見したら，そのテキストを編集して「アノテーション」ができる機能である．ここでのアノテーションは，ポッドキャストに対して書き起こしテキストを作成することを意味し，各認識誤りの箇所において，競合候補の中から正しい候補を選択するか，正しいテキストをタイプして訂正する．そのために，閲覧時の全文表示画面とは別に，音声に同期してスクロールする図2のような画面で前後の見通し良く効率的な訂正ができる機能を用意した．これは，以前我々が提案した「音声訂正」[14] に基づくインタフェースであり，単語グラフを圧縮したconfusion network（信頼度付き競合候補）を求めることで，候補表示を可能にした．

3.2 PodCastleの実装と一般公開

PodCastleのシステム構成図を図3に示す．**Webクローラ**はポッドキャストを収集してデータベース管理部へ登録する．そして，認識処理を繰り返している複数の**音声認識器**から**音声認識状態管理部**へリクエストがあると，次に認識すべきエピソードが引き渡される．音声認識器がその認識処理を終えると，認識結果は音声認識状態管理部を経てデータベース管理

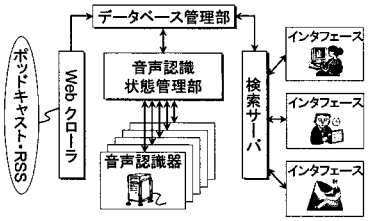

図 3: PodCastleのシステム構成図

部に渡される．**データベース管理部**では，ポッドキャストとその音声認識結果，ユーザによる訂正情報を索引付けして，処理状態の管理をする．最後に，**検索サーバ**は，Webサイトとしての機能を持ち，ユーザによる検索と**インタフェース**の画面遷移を管理する．

PodCastleは，http://podcastle.jp において2006年12月1日から一般公開され，それから約一ヶ月半の検索数は6151件であった．登録済のポッドキャスト数は117件，エピソード数は3820件で，そのうち一部でも訂正されたエピソード数は193件であった．

4 議論

音声認識研究2.0は，音声認識にWeb 2.0の考え方を導入したものであるが，以下では，その実例であるPodCastleがどの点でWeb 2.0的と言えるのかを考察し，様々な観点から今後の展望を議論する．

4.1 PodCastleはWeb 2.0に基づいているか

PodCastleというWebサービスと，Web 2.0との関連性を考察する．Web 2.0[6]は，Tim O'Reillyらによって2004年に提唱された概念で，近年，Web上の一連の新しい潮流を包含して説明する際に用いられることが多い．そこで以下では，インターネットアプリケーションであるPodCastleが持つ特長の中で，Web 2.0の考え方に基づいているものを列挙する．

- **集合知 (wisdom of crowds)，参加のアーキテクチャ，ユーザによる貢献**
  PodCastleは，Webの力を使って集合知を利用するというWeb 2.0の原則を実践している．不特定多数のユーザによる誤認識箇所の訂正が前提であり，そうしたユーザの参加を促すアーキテクチャを内在している．PodCastleは，ユーザの集合知によって，検索性能が改善していくポッドキャスト検索サービス，かつ，認識性能が改善していくポッドキャスト閲覧（半自動書き起こし）サービスと捉えることができる．そして，これらの改善がさらなるユーザの参加を促し，ユーザが増えるほど改善されるというソーシャルアノテーションのポジティブスパイラルが生まれる．
  ただし，ここで重要なのは「参加」つまり「訂正」の仕方と質である．音声認識性能によっては，ポッドキャストのエピソードを最初から最後まで訂正する作業は労力が大きく，数時間かかることがある．そこでPodCastleでは，そうした完全な訂正は求めずに，ユーザの気付いた範囲，可能な範囲の一部分だけでも訂正して貢献すると，性能が向上する仕組みになっていることが重要となる．つまり，少数の人から多大な貢献を期待するのではなく，多数の人から少しずつの貢献を期待する立場を取る．一方，訂正の質の問題については，4.2節 で改めて議論する．
- **ロングテール (long tail)**
  有名なポッドキャストと有名でなく通常は発見されにくいポッドキャストは対等なので，PodCastle上の検索結果として表示されることで聴取が促される．さらに，3 章でも述べたように，全文テキストを公開することで，Google等の外部の一般的なテキスト検索エンジンから検索されることも意図している．ユーザは新たなポッドキャストをリクエストして追加できるので，さらに検索対象が拡大し，豊かなテールを築いていける．
- **パーマリンク (permalink)**
  PodCastleでは，ポッドキャストやそれを構成する各エピソードのURLは，パーマリンクとしてユーザが外部利用できることを重視している．これによりある特定のポッドキャストについて言及したいときに，RSSやMP3ファイルのURLでなく，その全文テキストが見られるPodCastleのパーマリンクを利用できて便利である．

4.2 今後の展望

PodCastleは，Web 2.0の「永久にベータ版」という考え方に基づき公開を開始したため，今後も以下に述べるような様々なアイディアで拡張を続けていく予定である．

- **マッシュアップ (mashup)，フォークソノミー (folksonomy)**
  Web 2.0が持つ重要な概念の中でまだ対応していないものに，マッシュアップとフォークソノミーがある．マッシュアップ用の各種APIに関しては，PodCastle側で整備することを検討中である．フォークソノミーに関しては，各ポッドキャスト，エピソードに対するタギング（ユーザによる任意のキーワードでのラベル付け）への対応も検討しているが，各エピソードはパーマリンク化してあるため，タギングをサポートした他のソーシャルブックマーク用Webサービス等とマッシュアップした方が，ユーザの利便性が高い可能性もある．

- **RSS の配信**
  PodCastle上で特定の検索語を含むエピソードを検索できるだけでなく，その検索語を含むエピソード群を購読し続けることができるRSSも配信する予定である．ただし，RSSには様々なフォーマットがあり，すべてに対応することは難しい．そこで，PodCastleでは最小限のRSSを配信し，あとはユーザ側でマッシュアップしてもらうことを期待することとする☆4．
- **ユーザによる訂正の質（いたずら対策）**
  我々はWeb 2.0の「ユーザを信頼する」立場から，基本的にはユーザによる訂正の質は高いものと考えており，実際に公開後に集まった訂正結果の質は高い．しかし，もし仮にユーザが故意に不適切な訂正（いたずら）をした場合には問題になるため，その信頼性を音響的に評価する方法の研究も進めている．例えば，訂正結果の中で読みが判明する箇所に関して音響信号とのアラインメントを求め，その音響尤度が低すぎたら信頼性の低い訂正結果と判定する方法等を検討している．
- **動画中の音声への対応**
  音声のポッドキャストの動画版に相当するビデオポッドキャストもWeb上で普及しつつあるため，その音声トラックを抜き出して音声認識によりテキスト化することで，ポッドキャスト同様に検索の対象とする予定である．
- **個人的な書き起こし用インタフェースへの対応**
  インタビューや会議，講演等を書き起こす需要は高く，PodCastleはそのために有用だが，それらの音声データはポッドキャストとして公開できないことが多い．そこで，他のユーザには開示されないアクセス制限をかけるオプションを用意することを検討している．ここで重要なのは，その場合でも，訂正結果は全体の性能向上に寄与し，逆にその恩恵も受けられることである．

上記以外にも，音声認識性能の改善や他言語への対応等を検討している．

5 おわりに

本稿では，これまでの音声認識研究と相補関係にある「音声認識研究2.0」という新たな研究アプローチを提案し，その実例かつ音声認識のキラーアプリケーションとして，集合知を活用した音声情報検索用Webサービス「PodCastle」を実現した．本研究の学術的意義は，不特定多数のエンドユーザに音声認識誤りを訂正する協力をしてもらうことで，音声認識性能，音声情報検索性能をどこまで高くできるかを探求することにある．同時に，日本語ポッドキャスト検索のための世界初のWebサービスを公開して，エンドユーザの役に立つという社会的意義も持っている．

さらに本研究は，音声コーパスを事前に用意することが困難な状況で，どのようにすれば音声認識が役に立つかを明らかにする点でも意義がある．一般に，充分なコーパスが用意できれば音声認識技術は有用であるが，その整備は多大なコストと労力を要する上に，適用範囲が限定される問題があった．それに対して本研究では，誤認識も含めて全テキストを外部公開し，不特定多数のユーザの訂正によって「音声認識を育ててもらう」方針を取った．この場合，誤認識が多いために批判を受けるリスクはあるが，そうした現状をユーザと共有してはじめて，音声認識技術の真の普及と発展があると我々は考える．本研究により，ユーザの貢献を積極的に取り込んで音声認識実用化へ向けて研究する重要性と将来性が明らかになり，多くの研究者が取り組むことで，今後の音声認識・音声情報検索の研究分野に新たな展開を引き起こすことができればと願っている．

謝 辞

PodCastleのWebサーバとクライアントの実装を担当して頂いた有限会社ブラジル（代表取締役 上津 竜太郎 氏）と有限会社メロートーン（代表取締役 新井 俊一 氏）に感謝する．

参 考 文 献

[1] 嵯峨山茂樹: なぜ音声認識は使われないか・どうすれば使われるか？, 情処研報音声言語情報処理 94-SLP-1-4, 23–30 (1994).
[2] 中川聖一: 音声言語処理の進歩と今後, 情処研報 音声言語情報処理 2004-SLP-50-4, 23–30 (2004).
[3] 畑岡信夫: 音声技術実用化の課題と取り組み, 情処研報 音声言語情報処理2005-SLP-55-1, 1–6 (2005).
[4] 赤堀一郎, 渡辺隆夫, 河井恒, 庄境誠, 畑岡信夫: パネルディスカッション「音声認識技術の実用化」, 情処研報 音声言語情報処理2005-SLP-58-6, 31–40 (2005).
[5] 石川泰, 神沼充伸, 中川聖一, 磯健一, 新田恒雄: パネルディスカッション「音声認識の実用化の阻害要因と課題」, 情処研報音声言語情報処理 2006-SLP-63-9, 45–54 (2006).
[6] O'Reilly, T.: What Is Web 2.0 — Design Patterns and Business Models for the Next Generation of Software, http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.
[7] 李晃伸: 大語彙連続音声認識エンジンJuliusの開発の進展, 情処研報音声言語情報処理 2005-SLP-59-22, 127–132 (2005).
[8] 鹿野清宏, Tobias, C., 川波弘道, 西村竜一, 李晃伸: 音声情報案内システム「たけまるくん」および「キタちゃん」の開発, 情処研報音声言語情報処理 2006-SLP-63-7, 33–38 (2006).
[9] Wikipedia: http://www.wikipedia.org/.
[10] 緒方淳, 後藤真孝, 江渡浩一郎: PodCastle: ポッドキャストをテキストで検索, 閲覧, 編集できるソーシャルアノテーションシステム, WISS 2006 論文集, 53–58 (2006).
[11] Podscope: http://www.podscope.com/.
[12] PodZinger: http://www.podzinger.com/.
[13] 緒方淳, 後藤真孝, 江渡浩一郎: PodCastleの実現: Web 2.0に基づく音声認識性能の向上について, 情処研報 音声言語情報処理2007-SLP-65-8 (2007).
[14] 緒方淳, 後藤真孝: 音声訂正：選択操作による効率的な誤り訂正が可能な音声入力インタフェース, 情処学論, **48**, 1, 375–385 (2007).
[15] Plagger: http://plagger.org/.

☆4 例えば，Plagger[15] 等の外部のカスタマイズ可能なフィードアグリゲータで，最小限のRSSを利用して，任意の形式のRSSや付加情報を持つRSSを生成して使うことができる．