PodCastle

Speech Recognition Research 2.0

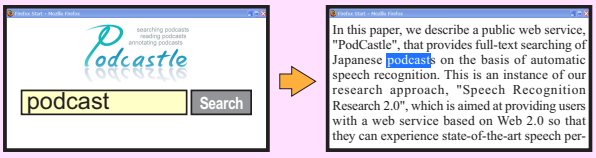
Motivation 1

❑ Users do not understand how useful ASR (automatic speech recognition) can be

- Researchers understand what sort of speech is easily recognized by ASR
- If users have previously had difficulty being understood by ASR, they doubt the usefulness and stop using it

➡ Promote the popularization and use of ASR by launching a web service *"PodCastle"*
for searching podcasts
for reading text of ASR results for podcasts

➡ Users do not have to provide their own speech input at all



What Are Podcasts?

❑ Audio programs distributed on the web (like radio shows or audio blogs)

- Podcast = RSS syndication feed + MP3 files

RSS syndication feed

Metadata
Title: CNN News Update
Description: The latest news happening in the U.S. and around the world.
Episode 1
Title: CNN News Update (8-21-2007 7 AM EDT)
MP3: http://rss.cnn.com/...08-21-07-7AM.mp3
Episode 2
Title: CNN News Update (8-21-2007 6 AM EDT)
MP3: http://rss.cnn.com/...08-21-07-6AM.mp3
Episode 3
Title: CNN News Update (8-21-2007 5 AM EDT)
MP3: http://rss.cnn.com/...08-21-07-5AM.mp3
Episode ...
(New episodes can be added at arbitrary intervals)

- With RSS, updated episodes are automatically downloaded

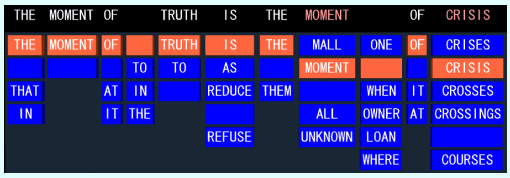
Motivation 2

❑ ASR cannot correctly transcribe podcasts

- Contents and recording conditions vary widely
- Preparation of corpora covering podcasts is too costly and time consuming

➡ Give up the idea of preparing corpora, and instead encourage users to cooperate
by correcting ASR errors to improve ASR/search performances

➡ Collaborative training for speech recognition



"Speech Repair" interface [Ogata & Goto, Interspeech 2005]

Video clip of PodCastle:
http://staff.aist.go.jp/m.goto/PodCastle/

Speech Recognition Research 2.0

❑ Definition

- Research approach where the current state of ASR is intentionally disclosed to users so that ASR performance can be improved through cooperative participation by users
- Named to reflect the concept of Web 2.0

❑ Goal

- Change the usage of ASR by setting the **positive spiral** into motion
- ASR-based web service that is permanently in beta version (***perpetual beta***) is launched and then improved by inviting users to use it on the web, thereby advancing the research

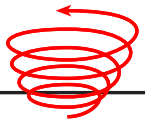
❑ PodCastle project

- Initiated in January 2006
- Japanese version was released to the public at *http://podcastle.jp* on December 1st, 2006

A Web 2.0 Approach to Speech Recognition Research

Speech Recognition Research 1.0	Speech Recognition Research 2.0
Stand-alone application	Web service
Dictation	Searching/browsing
Corpus	Web-based data
Limited topics	Unlimited topics
Transcription	Annotation
Out-of-vocabulary words	Not-yet-annotated words
Specialist participation	User participation
Individual correction	Social correction
Personal wisdom	Wisdom of crowds
Completed version	Perpetual beta

NOTE: We are not suggesting that Speech Recognition Research 1.0 (conventional approach, SRR-1.0) is inferior or obsolete. There is no doubt that continued research using the SRR-1.0 approach is needed. We ourselves have continued our work on SRR-1.0 as the foundation for 2.0. It should also be stressed that we are discussing research approaches, and not speech recognition techniques or algorithms themselves, which is why we use the term "Speech Recognition Research 2.0" instead of "Speech Recognition 2.0".



Positive spiral leading towards greater use of ASR

1. Allowing users to experience ASR lets them better understand its performance

✗ Once users experience ASR problems with their voices, they incorrectly assume that other people's voices will also not be well recognized

→ Promote understanding of ASR performance by providing a ***web service*** that allows users to ***search and browse*** open-to-the-public ***web-based speech data*** such as podcasts

2. Users contribute to improved ASR performance

✗ In-house improvements (voice adaptation and word registration) made by users are not made available for re-use by others

→ Enable recognition of various speech data on an ***unlimited range of topics*** by getting users to correct ASR errors
Users cooperate in the preparation of full-text transcriptions as a form of ***annotation***
User corrections are used for training ASR
Our-of-vocabulary words are regarded as being nothing more than ***not-yet-annotated words***

3. Improved performance leads to a better user experience

✗ Users have had little opportunity to experience the better performance that results from ongoing improvements made by researchers

→ Extend ***user participation*** framework to provide a ***social correction*** framework
Many anonymous users can improve ASR performance by sharing correction results and gain a real sense of contributing to the convenience of other users
Use the ***wisdom of crowds*** to achieve a better user experience

Masataka Goto, Jun Ogata, and Kouichirou Eto
National Institute of Advanced Industrial Sci. and Tech. (AIST)

Speech Recognition Research 2.0

Web 2.0 + Speech recognition + Podcast + Full-text speech retrieval + Wisdom of crowds

PodCastle

PodCastle

❑ Podcast search service based on ASR

- Users can search, read, and annotate podcasts
- Growing need for full-text speech retrieval service
- Existing podcast retrieval services (Podscope and EveryZing (PodZinger))
  - Hide full-text ASR results
  - Users have no means of correcting ASR errors
- PodCastle
  - Allow full-text ASR results to be accessed by both users and external search services
  - Allow users to cooperate with each other to improve ASR performance
- First instance of Speech Recognition Research 2.0

Three Functions

❑ Searching function

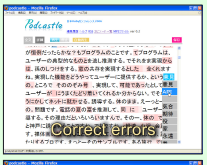
- Full-text search of ASR results
- List of episodes containing a search term is displayed together with text excerpts
- Each excerpt can be played back individually and be selected to read it

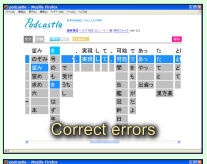
❑ Reading function

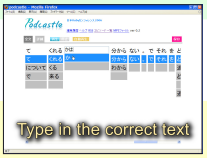
- View the full-text ASR result to understand the contents without audio playback
- Each word is colored according to the degree of ASR reliability
- Full text can be indexed and accessed by external search engines (e.g., Google)
  - Increase the value of podcasts by bringing more users into contact with them
  - Podcasters will be motivated to use the annotating function

❑ Annotating function (transcribing podcast contents)

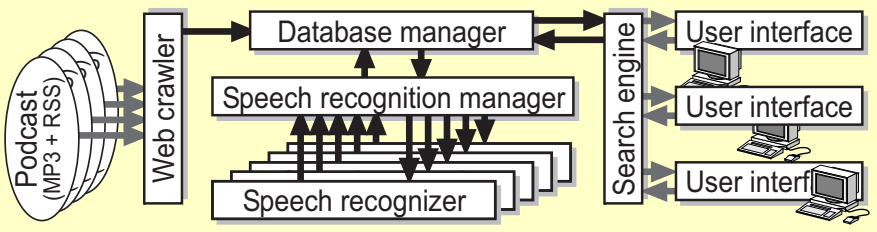
- Add *"annotations"* (transcription) to correct ASR errors
- Efficient error correction interface [Ogata & Goto, Interspeech 2005]
  - Select the correct candidate from the candidate list
  - Type in the correct text
- Candidate list is generated by using a *confusion network* that condenses a huge internal word graph of ASR







Implementation



JavaScript
Ruby on Rails
WEBrick
MySQL
Chasen
QuickTime/Flash
MTASC/MochiKit

Summary

❑ Research contribution
- Investigate how far the performance of ASR and full-text search can be improved by getting ASR errors corrected through cooperative efforts of many users

❑ Social contribution
- Help web users by providing the first public web service for full-text search of Japanese podcasts

❑ ASR contribution
- Demonstrate how ASR can be put to use in situations where a corpus is almost impossible to prepare

❑ Web 2.0 contribution (Original benefit not provided by Web 2.0)
- Automatic improvement: User contributions on a podcast can be automatically spread to other podcasts

❑ Our hope
- This study will prove the importance and potential of incorporating user contributions into ASR, and various other SRR-2.0-based projects will be done, thus adding a new dimension to this research field

See also: [Ogata, Goto, and Eto, Interspeech 2007]