PodCastle

Recent Advances of A Spoken Document Retrieval Service Improved by Anonymous User Contributions

Masataka Goto and Jun Ogata (AIST, Japan)

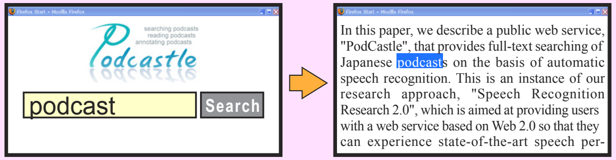
What is PodCastle?

searching podcasts
reading podcasts
annotating podcasts
Podcastle

Goal

❑ Full-text retrieval of speech data
- Podcasts (audio blogs)
- Individual audio files
- Video clips

(YouTube, Ustream.tv, and Nico Nico Douga)



In this paper, we describe a public web service, "PodCastle", that provides full-text searching of Japanese podcasts on the basis of automatic speech recognition. This is an instance of our research approach, "Speech Recognition Research 2.0", which is aimed at providing users with a web service based on Web 2.0 so that they can experience state-of-the-art speech per-

❑ ASR (automatic speech recognition) for text transcription
- Difficult to achieve high accuracy
- Diversity of topics, vocabularies, and speaking styles

Speech data
ASR result

The subprime loan crisis is far from over and ...

❑ Difficulties and Problems
- Cannot avoid making recognition errors for various types of speech data

Speech corpus cannot be prepared in advance
- Difficult to support new words/phrases (proper names and buzzwords)

Podcasts often include out-of-vocabulary words
- Difficult to launch a spoken document retrieval service with high accuracy

Users might be disappointed by ASR results

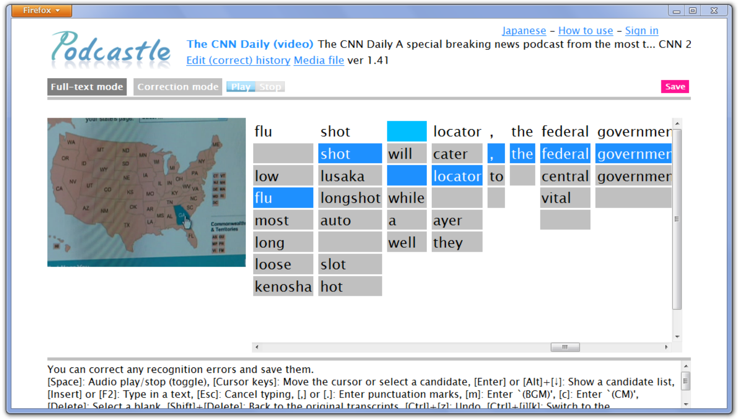
PodCastle

❑ Speech retrieval web service based on ASR and crowdsourcing
- Collect and amplify voluntary contributions by anonymous users

❑ Automatic learning from the web
- Automatically collect new words/phrases, their pronunciation, and usage examples

News articles (Yahoo! news) and web dictionaries
- Add new words to ASR dictionary (0.24M words)

❑ Users can find and correct ASR errors
- Original efficient error correction interface [Ogata & Goto, Interspeech 2005]
- Improve **retrieval performances** by correct indices
- Improve **recognition performances** by automatic learning (adaptation/training)



The CNN Daily (video) The CNN Daily A special breaking news podcast from the most t... CNN 2
ver 1.41

flu	shot		locator	,	the	federal	governmer
	shot	will	cater	,	the	federal	governmer
low	lusaka		locator	to		central	governmer
flu	longshot	while				vital	
most	auto	a	ayer				
long		well	they				
loose	slot						
kenosha	hot						

You can correct any recognition errors and save them.
[Space]: Audio play/stop (toggle), [Cursor keys]: Move the cursor or select a candidate, [Enter] or [Alt]+[↓]: Show a candidate list,
[Insert] or [F2]: Type in a text, [Esc]: Cancel typing, [,] or [.]: Enter punctuation marks, [m]: Enter `(BGM)', [c]: Enter `(CM)',
[Delete]: Select a blank, [Shift]+[Delete]: Back to the original transcripts, [Ctrl]+[z]: Undo, [Ctrl]+[i][k]: Switch to the

Three Functions

❑ Searching function
- Full-text search of ASR results
- List of speech data containing a query is displayed together with text excerpts
- Excerpts can be played back individually

❑ Reading function
- View the transcribed text of speech data
- 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)

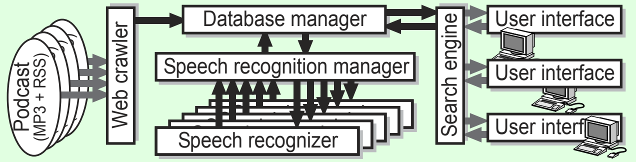
❑ Annotating function (error correction)
- Add *"annotations"* to correct ASR errors
- Select the correct candidate from the list

The list is generated by using a confusion network that condenses a huge internal word graph
- Type in the correct text
- Corrected errors can be used for improving retrieval and recognition performances

Implementation

JavaScript
Ruby on Rails
MySQL/Senna

Chasen
QuickTime/Flash
MTASC/Flex



Recent Advances

History

❑ http://podcastle.jp since 2006

- 2006/01 Started the project
- 2006/12 Released to the public

  The world's first speech retrieval project using crowdsourcing
- 2007/08 Interspeech 2007 papers

  Speech Recognition Research 2.0
- 2008/06 Press release

  Reported in TV/web news, newspapers, etc.
- 2009/08 Supported *video podcasts*
- 2009/09 Interspeech 2009 paper
- 2011/01 Supported *YouTube/Ustream.tv*
- 2011/03 Supported *Nico Nico Douga*
- 2011/?? Launch the English version

❑ Recently supported functions

- Support video sharing services
- Annotate speaker names and paragraphs
- Mark (change the color of) correct words that do not need any correction
- Show the percentage of correction

  100% when all words are corrected or marked
- Support simultaneous correction by users

  Corrections can be automatically shared (synchronized) and shown on their screens

  Useful for serious and rapid transcription

Experiences

❑ How widely used? (as of May 31st, 2011)

- 765 Japanese speech programs

  Podcasts and YouTube channels
- Consist of 112,476 audio files in total
- 2,593 audio files were partially corrected
- 521,938 corrected words (errors)

  52.8% were corrected by the candidate selection

  47.2% were corrected by the text typing
- There are users who voluntarily cooperate in the correction

  Speech data recorded by famous artists and TV personalities tend to receive many corrections

  Some podcasts were corrected almost everyday or every week

❑ ASR performance improvements

- Collaborative training of speech recognizer
- Podcast-dependent acoustic model trained using transcripts corrected by users

  [Ogata & Goto, Interspeech 2007, 2009, SSCS 2009]
- Confirmed that ASR performance for podcasts receiving many corrections was actually improved by this AM training

  Relative error reduction of 21-33%

  [Ogata & Goto, Interspeech 2009]
- Confirmed that ASR performance was also improved by language model training

Motivations

❑ Why did users correct errors?

- Correction itself is enjoyable and interesting
- Users want to contribute
- Users want their speech data to be correctly searched
- Users like the content and cannot tolerate the presence of recognition errors in it

Summary

❑ Technical contribution

- Investigate how far the ASR performance can be improved

  through the cooperative efforts of many end users
- PodCastle: Social correction framework

  Users gain a real sense of contributing to the convenience of themselves and other users
- Other game-based approaches often depend on the feeling of fun

  Human Computation or GWAPs (games with a purpose)

  Lack the feeling that the improved performance leads to a better user experience

❑ ASR contribution

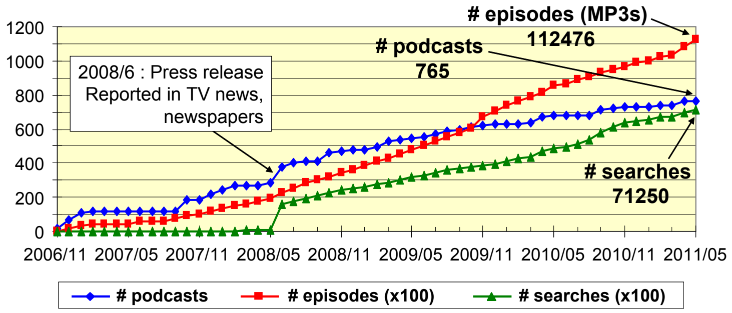
- Demonstrate how ASR can be put to use

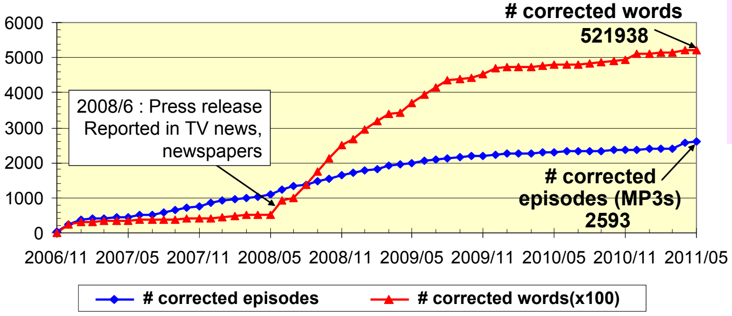
  in situations where a corpus is difficult to prepare

❑ Beyond Web 2.0 and Human Computation

- Framework for amplifying user contributions

  Improvements are automatically spread to other items not contributed by users





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

2011/08/28 Interspeech 2011 poster