PodCastle and Songle: Web Services for Retrieval and Browsing of Speech and Music Content on the Basis of Automatic Content Analysis and Crowdsourcing

AIST (National Institute of Advanced Industrial Science and Technology)

Masataka Goto

Jun Ogata, Kazuyoshi Yoshii, Hiromasa Fujihara, Matthias Mauch, Tomoyasu Nakano

2012/09/26 International Workshop on Search Computing
Toward Advanced Search Computing

Media Search + Web +
Content Analysis +
Visualization + HCl +
Crowdsourcing

= ∞
Self-Introduction

Masataka Goto

- Prime Senior Researcher / Leader of the Media Interaction Group, AIST
- 42 years old
- Working on
  - Music Information Research since 1992
  - Speech Interface Research since 1998
Self-Introduction

- Masataka Goto
  - Prime Senior Researcher / Leader of the Media Interaction Group, AIST
  - 42 years old
  - Working on Music Information Research since 1992
    Speech Interface Research since 1998

No. 4
### Music Information Research by M. Goto

<table>
<thead>
<tr>
<th>Year</th>
<th>Project/Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992</td>
<td>Drum transcription</td>
</tr>
<tr>
<td>1993</td>
<td>Cindy (Beat tracking system)</td>
</tr>
<tr>
<td>1994</td>
<td>VirJa Session (Virtual jazz session system)</td>
</tr>
<tr>
<td>1995</td>
<td>PreFEst (Melody/bass)</td>
</tr>
<tr>
<td>1996</td>
<td>SmartMusicKIOSK (Chorus detection)</td>
</tr>
<tr>
<td>1997</td>
<td>Musiccream &amp; MusicRainbow (Music listening interfaces)</td>
</tr>
<tr>
<td>1998</td>
<td>Drumix (Drum-sound recognition)</td>
</tr>
<tr>
<td>1999</td>
<td>LyricSynchronizer (Song-lyrics synchronization)</td>
</tr>
<tr>
<td>2000</td>
<td>VocalFinder (Content-based MIR)</td>
</tr>
<tr>
<td>2001</td>
<td>INTER (Score-informed separation)</td>
</tr>
<tr>
<td>2002</td>
<td>SingBySpeaking (Speech-to-singing)</td>
</tr>
<tr>
<td>2003</td>
<td>VocalListener &amp; VocalListener2</td>
</tr>
<tr>
<td>2004</td>
<td>VocaWatcher (Robot singer HRP-4C)</td>
</tr>
<tr>
<td>2005</td>
<td>Songle (Web service for active music listening)</td>
</tr>
<tr>
<td>2006</td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
</tr>
</tbody>
</table>

- **Active Music Listening Interfaces**
- **Music Understanding**
  - Music Interfaces
  - Singing Information Processing

- **Waseda Univ.**
- **AIST**

- **RWC Music Database**
Music Information Research by M. Goto

Waseda Univ. AIST

Drum transcription

Cindy (Beat tracking system)

VirJa Session (Virtual jazz session system)

PreFEst (Melody/bass)

SmartMusicKIOSK (Chorus detection)

Musicream & MusicRainbow (Music listening interfaces)

Drumix (Drum-sound recognition)

LyricSynchronizer (Songle)

VocalFinder (Content-based MIR)

Inter (Score-informed separation)

VocalListener & VocalListener2

VocaWatcher & VocaWatcher2

VocaWatcher (Robot singer HRP-4C)

Songle (Web service for active music listening)

RWC Music Database

Active Music Listening Interfaces

Music Understanding

Music Interfaces

Singing Information Processing

Music Information Research by M. Goto

Waseda Univ.  AIST


Drum transcription
Cindy (Beat tracking system)
VirJa Session (Virtual jazz session system)
PreFEst (Melody/bass)
SmartMusicKIOSK (Chorus detection)
Musiccream & MusicRainbow (Music listening interfaces)
Drumix (Drum-sound recognition)
Musicream & MusicRainbow (Music listening interfaces)

Active Music Listening Interfaces

SmartMusicKIOSK: Music Listening Interface

Music Understanding +
Music Interfaces +
Singing Information Processing

RWC Music Database

[Gotó, 2002-]
Music Information Research by M. Goto

Waseda Univ. AIST


Drum transcription

Cindy (Beat tracking system)

VirJa Session (Virtual jazz session system)

PreFEst (Melody/bass)

SmartMusicKIOSK (Chorus detection)

Musiccream & MusicRainbow (Music listening interfaces)

Drumix (Drum-sound recognition)

LyricSynchronizer (Singing Information Processing)

Musiccream: Music Discovery Interface

Music Understanding

+ Music Interfaces

+ Singing Information Processing

RWC Music Database

Active Music Listening Interfaces

[Goto, Goto, 2004]
Music Information Research by M. Goto

Waseda Univ. AIST

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Drum transcription</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cindy (Beat tracking system)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VirJa Session (Virtual jazz session system)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PreFEst (Melody/bass)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SmartMusicKIOSK (Chorus detection)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Musiccream &amp; MusicRainbow (Music listening interfaces)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drumix (Drum-sound recognition)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LyricSynchronizer (Song-lyrics synchronization)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VocalFinder (Choral singing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTER (Score-informed separation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SingBySpeaking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VocaListener &amp; VocaListener2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Music Understanding + Music Interfaces + Singing Information Processing

Songle (Web service for active music listening)

RWC Music Database

Active Music Listening Interfaces
Self-Introduction

Masataka Goto

- Prime Senior Researcher / Leader of the Media Interaction Group, AIST
  National Institute of Advanced Industrial Science and Technology

- 42 years old
- Working on

Music Information Research since 1992
Speech Interface Research since 1998
Speech Interface Research by M. Goto

Waseda Univ. AIST


Filled-pause detection
Speech Completion
Speech Shift
Speech Starter
Speech Spotter
Speech Repair
Speech Pen
PodCastle (Web service for speech retrieval)
Speech Interface Research by M. Goto

Waseda Univ. AIST


Filled-pause detection
Speech Completion
Speech Shift
Speech Starter
Speech Spotter
Speech Repair
Speech Pen

PodCastle (Web service for speech retrieval)

[Goto et al., 2000-]
Speech Interface Research by M. Goto

Filled-pause detection
Speech Completion
Speech Shift
Speech Starter
Speech Spotter
Speech Repair
Speech Pen

PodCastle (Web service for speech retrieval)
PodCastle and Songle: Web Services for Retrieval and Browsing of Speech and Music Content on the Basis of Automatic Content Analysis and Crowdsourcing

AIST (National Institute of Advanced Industrial Science and Technology)

Masataka Goto

Jun Ogata, Kazuyoshi Yoshii, Hiromasa Fujihara, Matthias Mauch, Tomoyasu Nakano

2012/09/26 International Workshop on Search Computing
Our Goal

- Provide users with public web services so that they can experience state-of-the-art research-level technologies
  - Speech recognition, music understanding, signal processing, machine learning, and crowdsourcing
- Meet growing needs for retrieval of media content
- Toward this goal, we developed two web services for content-based retrieval and browsing
  - PodCastle for speech
  - Songle for music
Our Goal

- Provide users with public web services so that they can experience state-of-the-art research-level technologies.
  - Speech recognition, music understanding, signal processing.

Toward this goal, we developed two web services:
- PodCastle for speech
- Songle for music

Originality

We collect and amplify voluntary contributions by anonymous users to improve user experiences.
PodCastle for speech
PodCastle (http://podcastle.jp)

- Spoken content retrieval service based on automatic speech recognition (ASR)
- Provide full-text searching of speech data
  - Video clips on video sharing services (YouTube, Ustream, etc.)
  - Podcasts
  - Individual audio or movie files on the web

PodCastle enables a user to

1) find speech data that include a search term
2) read full texts of their recognition results
3) easily correct recognition errors
Performance Improvement by Users

- Anonymous users can **find and correct** recognition errors

- **Novel efficient error correction interface** [Ogata, Goto, 2004]
  - Select the **correct candidate** from the candidate list
Performance Improvement by Users

- Anonymous users can find and correct recognition errors

- Novel efficient error correction interface [Ogata, Goto, 2004]
  - Select the correct candidate from the candidate list

Corrected errors can be used for

- Improving retrieval performances by correct indices

- Improving recognition performances by machine learning (adaptation/training)
Performance Improvement by Users

- Anonymous users can find and correct errors
- Novel efficient error correction interface
  - Select the correct candidate from the candidate list

Corrected errors can be used for
- Improving retrieval performances by correct indices
- Improving recognition performances by machine learning (adaptation/training)

[Ogata, Goto, 2004]
History

- **PodCastle Project** (http://en.podcastle.jp)
  - 2006/01  Started the project
  - 2006/12  Released to the public
    - The world’s first speech retrieval using crowdsourcing
  - 2008/06  Press release
    - Released to the public
    - Reported in TV news, newspapers, web news, etc.
History

- **PodCastle Project** (http://en.podcastle.jp)
  - 2006/01 Started the project
  - 2006/12 Released to the public
    - The world’s first speech retrieval using crowdsourcing
  - 2008/06 Press release
    - Released to the public
    - Reported in TV news, newspapers, web news, etc.
  - 2009/08 Started supporting video podcasts
  - 2011/10 Press release
    - Launched the English version (CSTR, U. Edinburgh)
    - Started supporting video sharing services
How Widely Used Is PodCastle?

- **Size of data stored in PodCastle**

  - # podcasts: 914
  - # episodes (audio/video files): 188,538
  - # searches: 134,319

  **Reported in TV/web news, newspapers, etc.:**
  - 2008/6, 2011/10: Press release
How Widely Used Is PodCastle?

- Amount of corrections

- # corrected episodes
- # corrected words (x100)

2008/6, 2011/10: Press release
Reported in TV/web news, newspapers, etc.

# corrected words
590579

# corrected episodes
3598
How Widely Used Is PodCastle?

- As of September 1, 2012
  - # registered channels: 914
  - # registered episodes (MP3s): 188,538
  - # corrected episodes (MP3s): 3,598
  - # corrected words: 590,579

- Some channels (podcasts) have been corrected almost everyday or every week

- Content by famous artists and TV personalities tend to receive many corrections
Performance Improvement

- Reduce recognition errors

Before learning

Corrected errors

After learning

Corrected errors

Errors (blue words) were reduced
Why Do Users Correct Errors?

- Inferred motivations

  1. Error correction itself is enjoyable and interesting
     Using our interface could be fun
     somewhat like the fun in a video game

  2. Users want to contribute
     Some users correct errors not only for their own convenience,
     but also to altruistically contribute to
     the improvement of speech recognition and retrieval
Why Do Users Correct Errors?

- Inferred motivations

3. Users want their speech data to be correctly searched
   Creators of speech data would correct errors so that their speech data can be searched more accurately

4. Users like the content and cannot tolerate the presence of recognition errors
   Some fans of famous artists or TV personalities would correct errors because they like the speakers' voices and cannot tolerate the presence of errors in their favorite content
Discussion

.framework for amplifying user contributions

• Typical Web 2.0 services
  Improvements are limited to an item directly contributed by users

• Beyond Web 2.0: Advantage of PodCastle
  Improvements are automatically spread to other items not contributed by users
  New technology of amplifying user contributions to improve performances

(cf. Human computation (ESP Game), GWAP (Game With A Purpose) by Luis von Ahn)
Summary of PodCastle

- **The users are happy**
  - because our web service is getting better and better

- **We are happy**
  - because our ASR system is getting better and better

- **We would eventually obtain the best ASR system in the world**
Songle for music
Songle (http://songle.jp)

- Web service for active music listening
  - Allow anonymous users to enjoy any songs (MP3 files) available on the web by using active music listening interfaces
Use automatic music-understanding technologies

- Estimate four major types of music scene descriptions

- A user can enjoy playing back a song while seeing the visualization of the estimated descriptions

**Music structure**
(chorus / repeated sections)

**Chords**
(root note and chord type)

**Melody line**
(F0 of the vocal melody)

**Beat structure**
(musical beats and bar lines)
Demonstration (Visualization)
Songle facilitates deeper understanding of music by visualizing music scene descriptions.
Songle with SmartMusicKIOSK

- Implement all functions of SmartMusicKIOSK
  - Jump and listen to the **chorus** with just a push of a button

[Goto, ACM UIST 2003]
Songle with SmartMusicKIOSK

- Implement all functions of SmartMusicKIOSK
  - Jump and listen to the *chorus* with just a push of a button

[Songle makes it easier for a user to find desired parts of a piece](Goto, ACM UIST 2003)
New Function: Visualizer

- Visualizer Mode
  - Four types of animation rigidly synchronized with music
New Function: Content-based Retrieval

- Chord Progression Search
  - Listen to songs having the same sequence of chord names
## New Function: Songle Embedded Player

- **Songle Embedded Player**
  - Embed not only **Songle Embedded Player** but also **music-sync animation** in your homepage or blog.

---

> A web service for active music listening "Songle" ([http://songle.jp](http://songle.jp)) enables you to embed the Songle Embedded Player in your homepage or blog.

The music structure of a song is shown in the center of the player. If you push an orange region or a right orange button, you can immediately listen to the chorus.

In addition, you can show **music-synced animation** in the background of your web page. This page is intended to demonstrate this function.

Note that the music-synced animation cannot be shown on Internet Explorer 8 or below because HTML5 is not supported. **Chrome browser** is recommended.

This function became available after **AIST’s press release** on August 29th, 2012.

### Examples of web pages demonstrating the Songle Embedded Player:
- Miky 5th Anniversary
- CIPRO blog

### How to embed the Songle Embedded Player:
Anonymous users can **find and correct** estimation errors.

**Efficient error correction interface (editor)**
- Easily correct the error by selecting from a list of candidates or by providing an alternative description.
Demonstration (Error Correction)

- Music structure
Demonstration (Error Correction)

- Beat structure

![Diagram of Beat structure](image)
Demonstration (Error Correction)

- Melody line
Demonstration (Error Correction)

- Chords
Demonstration (Error Correction)

- Chords

Corrected errors can be shared and used to immediately improve the user experience
Let’s Enrich Music Listening Experiences

- **Songle** is for better user experiences, not just for collecting annotations
  - Users can simply enjoy **active music listening** without **correcting** any errors!
  - We understand that it is too difficult for some users to **correct**

- **Users are not expected to correct all errors**, **only some** according to each user's interests
History

- **Songle Project**
  - 2011/10  Launched the alpha version
  - 2012/02  Released the beta version
  - 2012/08  Press release

Released to the public
Reported in Newspapers, web news, etc.

http://songle.jp

English and Japanese versions are available.
You can register any song (MP3 file) on the web!
Collaboration for End Users

- Collaboration with PIAPRO (CGM content site)
  CGM = Consumer Generated Media (Crowdsourced Content)
  - 2012/08 Started official collaboration
    Web site for creators to post CGM content and collaborate

Link to Songle from each song of PIAPRO (80000 songs)

- Music analysis and correction
- Chorus-search function on PIAPRO site
Conclusion

PodCastle for speech

Songle for music
PodCastle and Songle serve as a *showcase*

- Make **social contributions** by providing public web services that let people **retrieve speech data** and let people **enjoy active music listening interfaces**

- Promote the **popularization and use of speech-recognition and music-understanding technologies** by raising user awareness. Demonstrate how people can **benefit from these technologies**

- Users can grasp their **nature** through **user experiences**
  
  *“What kinds of speech and music are difficult to handle?”*
Social Contributions

- **PodCastle and Songle serve as a showcase**
  - Prevent overestimation of the technologies behind
  - The originally estimated values are visualized as trails with different colors after user corrections

- All correction histories are recorded and can be compared
PodCastle and Songle

- New research approach to speech recognition and music understanding
- Designed to set into motion a positive spiral

1. We enable users to experience a service based on our technology to let them better understand its performance.
2. Users contribute to improved performance.
3. The improved performance leads to a better user experience, encouraging further use of the service at step (1).

Better user experiences

Game-based or crowdsourcing approaches often lack step (3) and depend on the feeling of fun or money.

[Turnbull et al., '07] [Mandel et al., '07] [Law et al., '07] [Lee, '10] [Mandel et al., '10]
Academic Contributions

- **PodCastle/Songle as Social Correction Framework**
  - Users gain a real sense of contributing for their own benefit and that of others
  - Users can be further motivated to contribute by seeing corrections made by other users

- **Future work**
  - Use corrected errors and machine learning techniques to improve music-understanding technologies for Songle
  - Better user experiences
  - Better technologies
Let’s Enjoy!


English and Japanese versions are available
Toward Advanced Search Computing

Media Search + Web + Content Analysis + Visualization + HCI + Crowdsourcing = ∞
Acknowledgments

- Tomoyasu Nakano (for VocaListener and VocaWatcher)
- Shuuji Kajita, Yosuke Matsusaka, Shin'ichiro Nakaoka, and Kazuhito Yokoi (for VocaWatcher)
- Takayuki Goto (for Musicream)
- Katunobu Itou and Satoru Hayamizu (for Speech Completion)
- JST OngaCREST Project (for the current research funding)
  PI: Masataka Goto [2011-2017]

Please send me your comments:

E-mail: m.goto@aist.go.jp
URL: http://staff.aist.go.jp/m.goto/