

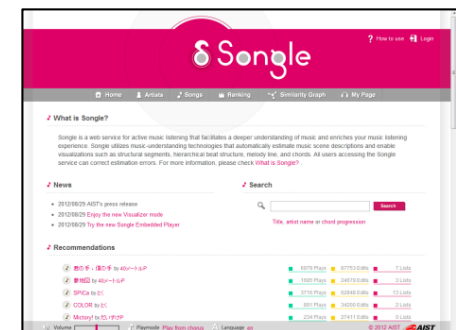
PodCastle and Songle:

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**





Toward Advanced Search Computing

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

Self-Introduction

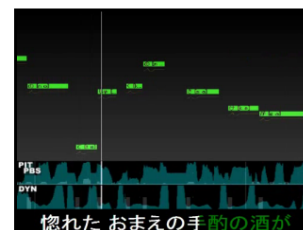
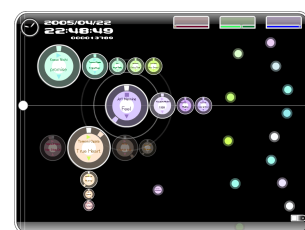
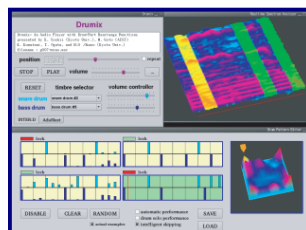
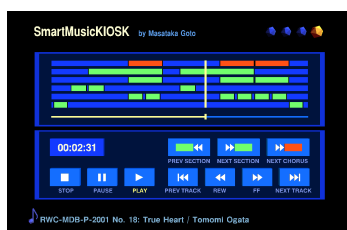


□ Masataka Goto

- 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



Self-Introduction

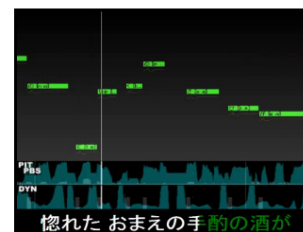
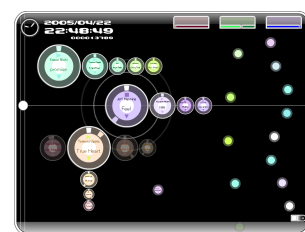
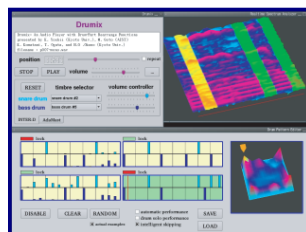
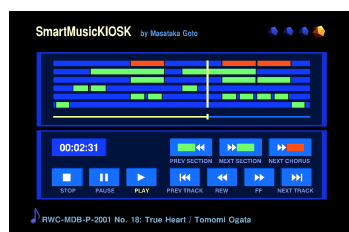


□ Masataka Goto

- 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



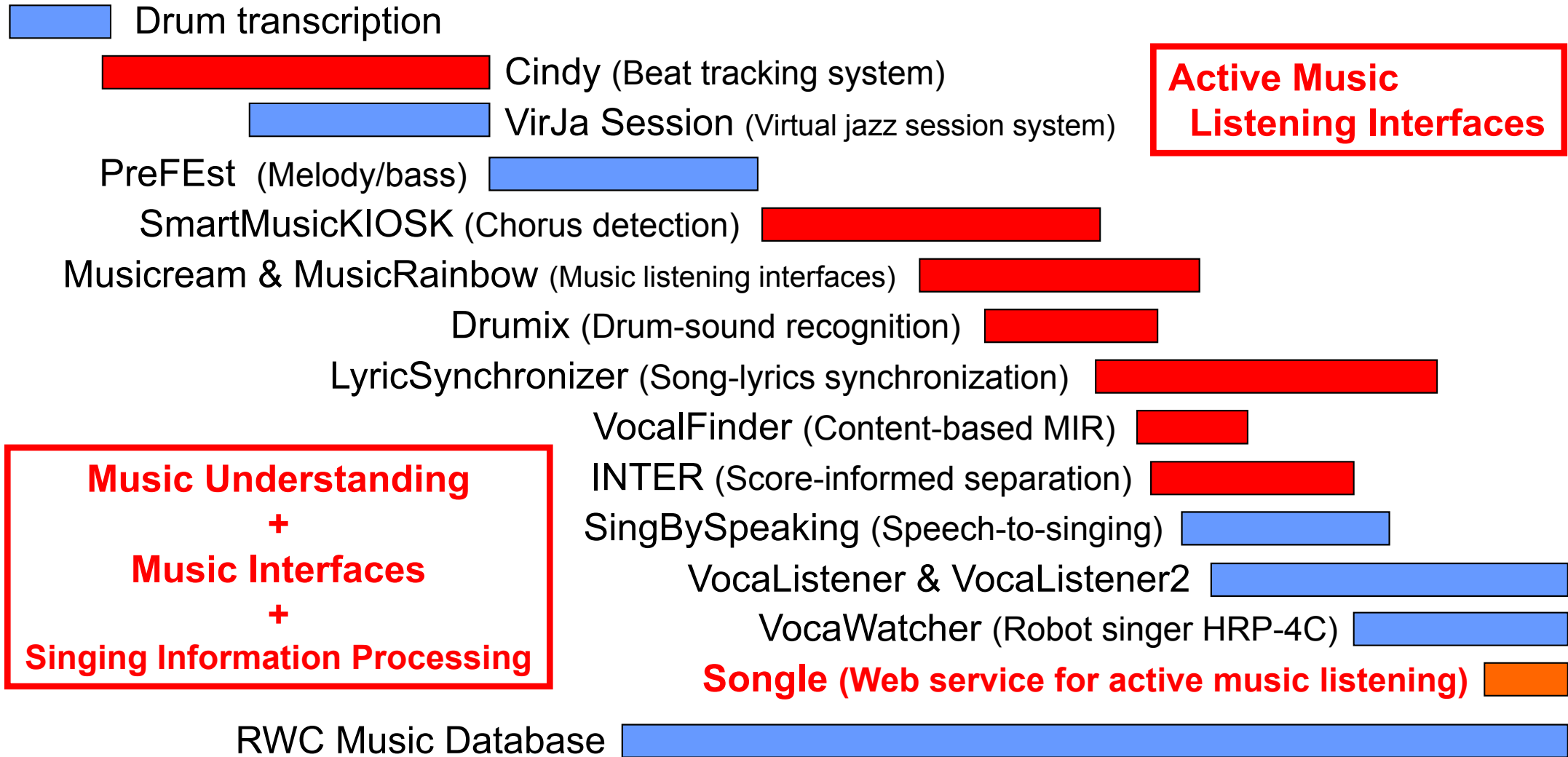


Music Information Research by M. Goto

Waseda Univ.

AIST

1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------





Music Information Research by M. Goto

Waseda Univ.

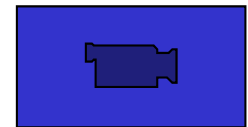
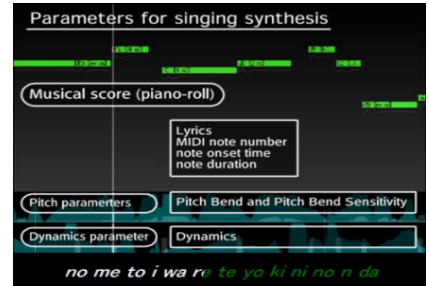
AIST

1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

- Drum transcription
- Cindy (Beat tracking system)
- VirJa S
- PreFEst (Melody/bass)
- SmartMusicKIOSK (Chorus d
- Musicream & MusicRainbow (Mus
- Drumix (D
- LyricSynchronize
- Voc

Active Music

VocaListener (Singing synthesis)



Packaged (Megpoid)

[Nakano, Goto, 2008-]

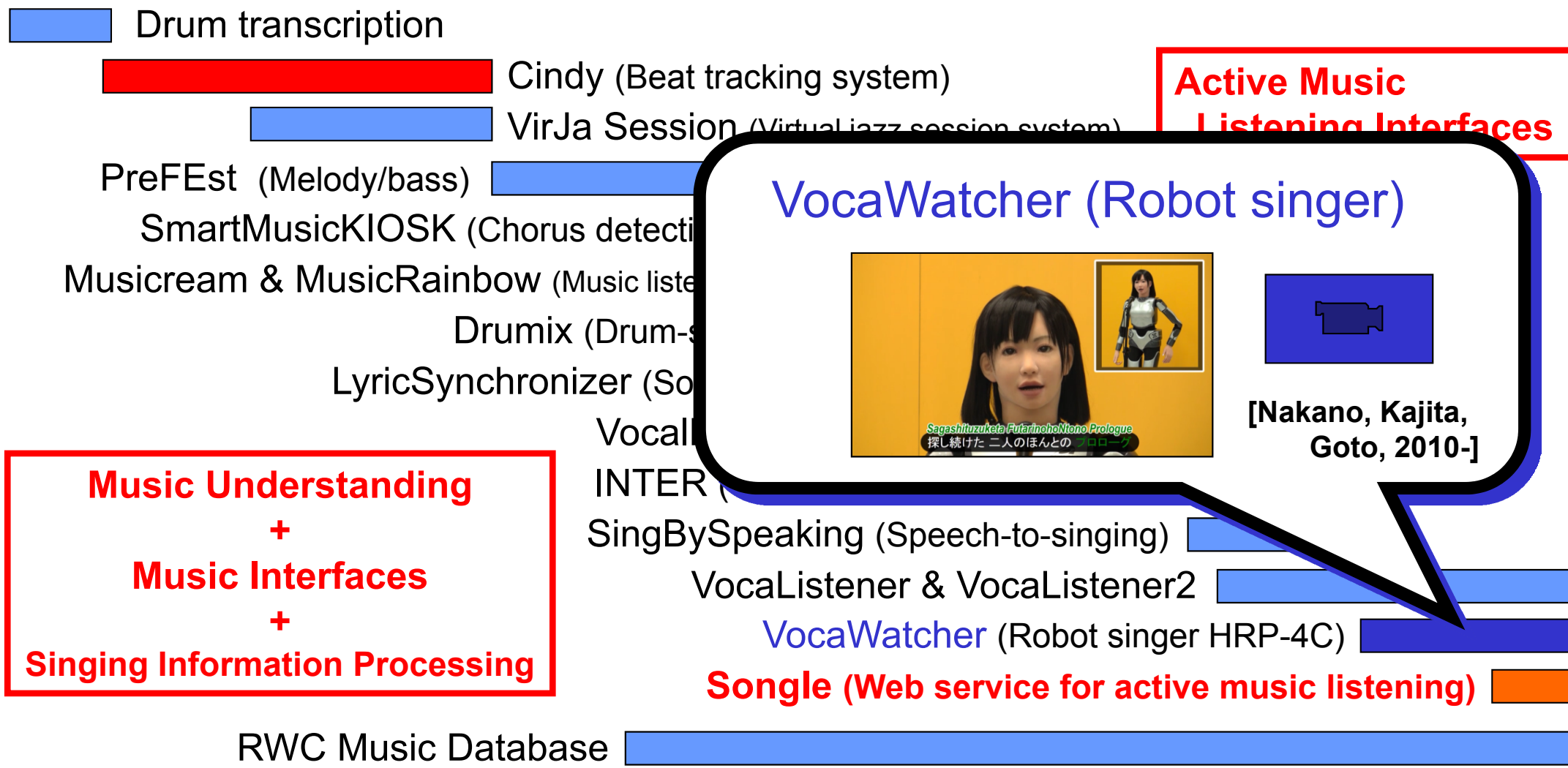
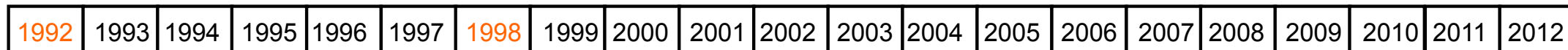
Music Understanding
+
Music Interfaces
+
Singing Information Processing

- INTER (Score-informed separation)
- SingBySpeaking (Speech-to-singing)
- VocaListener & VocaListener2
- VocaWatcher (Robot singer HRP-4C)
- Songle (Web service for active music listening)
- RWC Music Database

Music Information Research by M. Goto

Waseda Univ.

AIST



Active Music Listening Interfaces

VocaWatcher (Robot singer)

[Nakano, Kajita, Goto, 2010-]

Music Understanding
+
Music Interfaces
+
Singing Information Processing

Songle (Web service for active music listening)



Music Information Research by M. Goto

Waseda Univ.

AIST

1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

Drum transcription

Cindy (Beat tracking system)

VirJa Session (Virtual jazz session system)

PreFEst (Melody/bass)

SmartMusicKIOSK (Chorus detection)

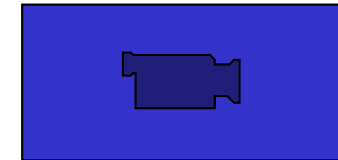
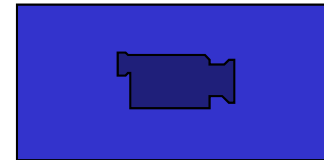
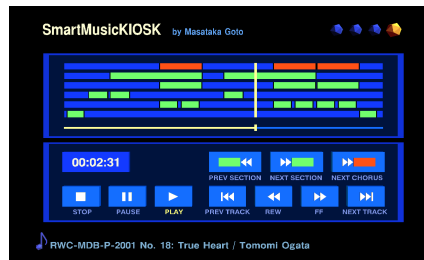
Musiccream & MusicRainbow (Music listening)

Drum transcription

**Active Music
Listening Interfaces**

SmartMusicKIOSK: Music Listening Interface

Music Understanding
+
Music Interaction
+
Singing Information



[Goto, 2002-]

RWC Music Database





Music Information Research by M. Goto

Waseda Univ.

AIST

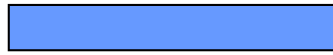
1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------



Drum transcription



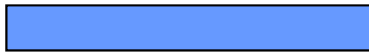
Cindy (Beat tracking system)



VirJa Session (Virtual jazz session system)

**Active Music
Listening Interfaces**

PreFEst (Melody/bass)



SmartMusicKIOSK (Chorus detection)



Musicream & MusicRainbow (Music listening interfaces)



Drumix (Drum-sound recognition)

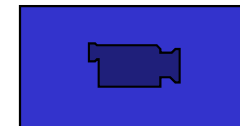
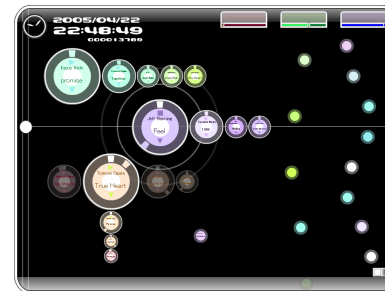


LyricSynchronizer (Song synchronization)



Music Understanding
+
Music Interfaces
+
Singing Information Processing

Musicream: Music Discovery Interface



[Goto, Goto, 2004]

RWC Music Database

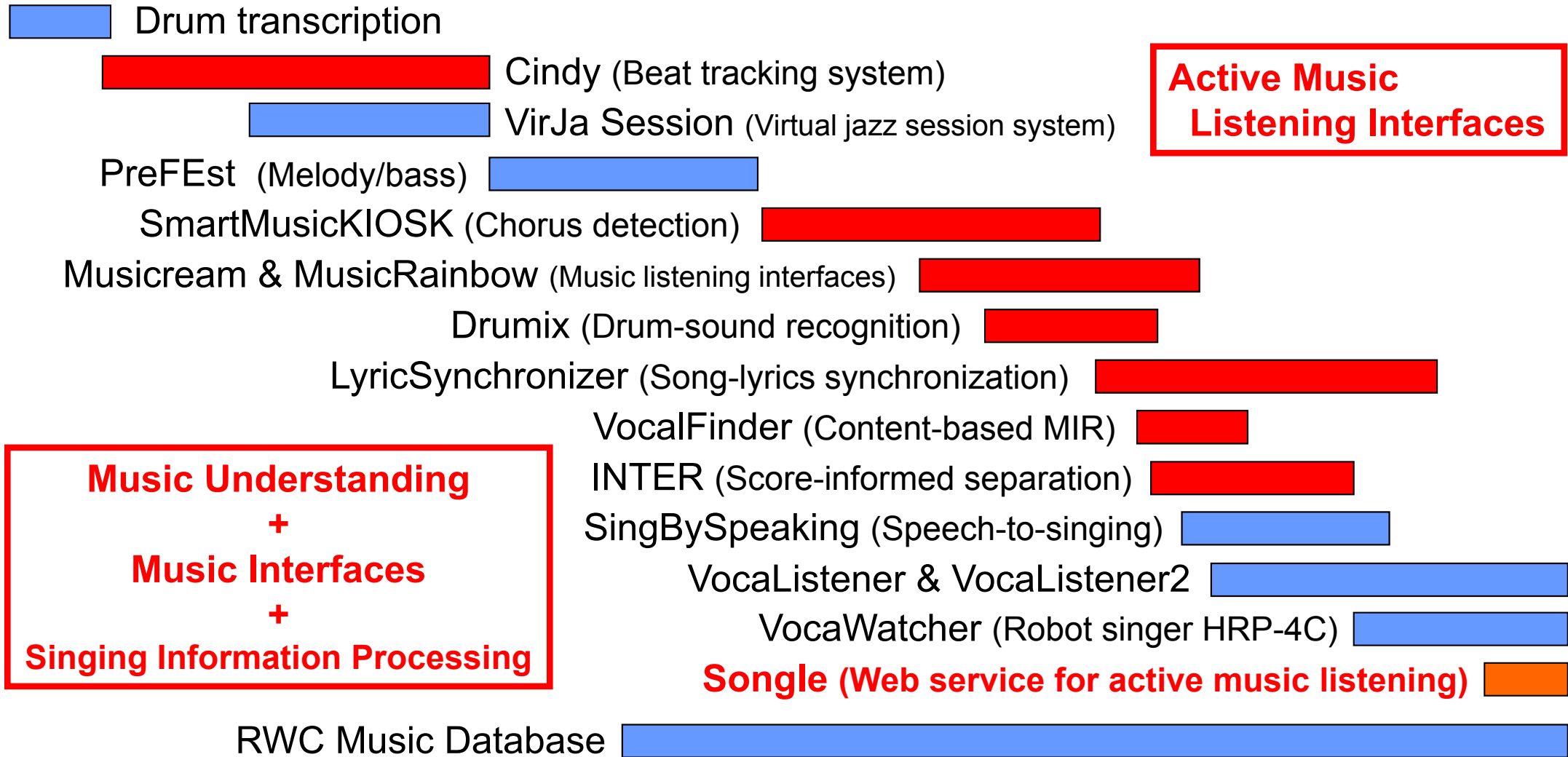


Music Information Research by M. Goto

Waseda Univ.

AIST

1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------





Music Information Research by M. Goto

Waseda Univ.

AIST

1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

Drum transcription

Cindy (Beat tracking system)

VirJa Session (Virtual jazz session system)

Active Music Listening Interfaces

PreFEst (Melody/bass)

SmartMusicKIOSK (Chorus detection)

Musicream & MusicRainbow (Music listening interface)

Drumix (Drum-sound recognition)

LyricSynchronizer (Song-lyrics synchronization)

VocalFinder (Chorus detection)

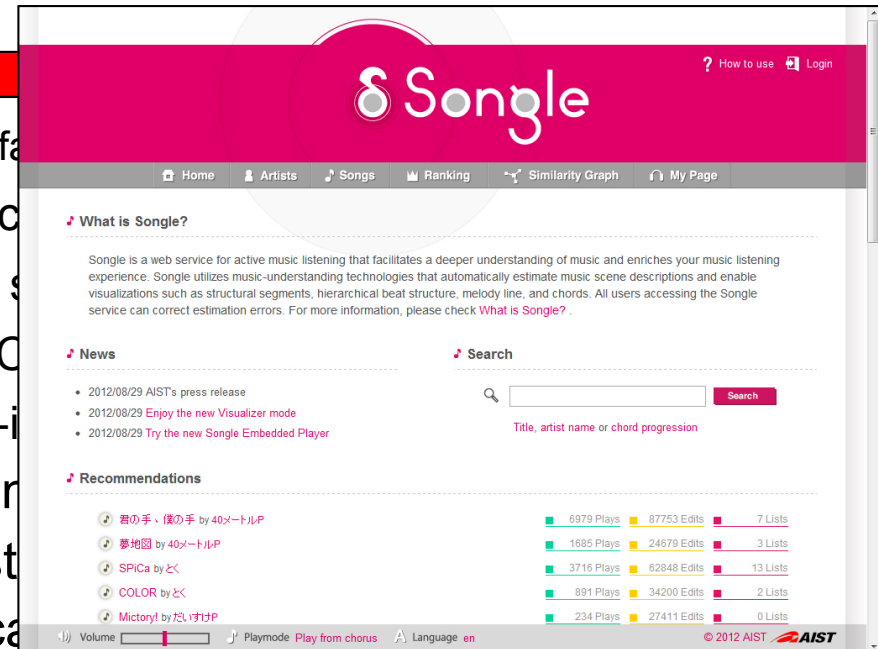
INTER (Score-independent)

SingBySpeaking

VocaList

Voca

Music Understanding
+
Music Interfaces
+
Singing Information Processing



Songle (Web service for active music listening)

RWC Music Database

Self-Introduction



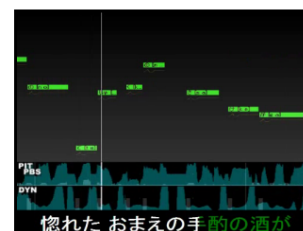
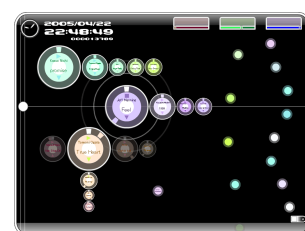
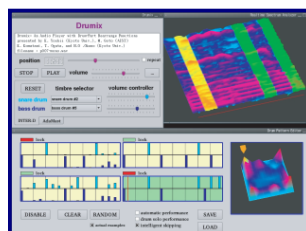
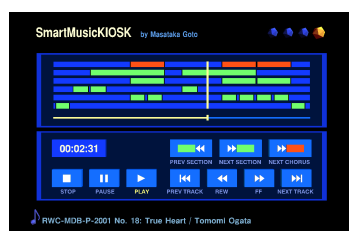
□ Masataka Goto

- 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

1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

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

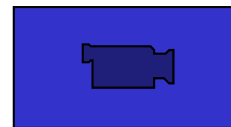
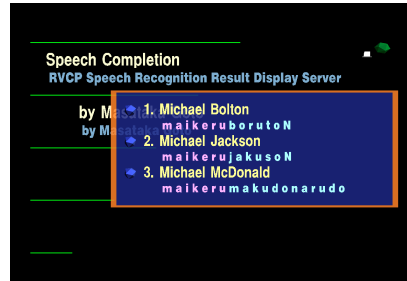
1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

Filled-pause detection 

Speech Completion 

Speech 

Speech Completion



[Goto et al., 2000-]





Speech Interface Research by M. Goto

Waseda Univ.

AIST

1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

Filled-pause detection



Speech Completion



Speech Shift

Speech S

Speed

S

PodCastle

searching podcasts
reading podcasts
annotating podcasts

Search Advanced search

Recommended Queries: [bit our information times called problem things talking](#)

Recommend

- [AIST - Paving the way to the Future](#) Media files in www.youtube.com **260 Edits** **28.4%** Video **1 fav**
- [CrestMuse: demo video \(4/5\)](#) Media files in www.youtube.com **142 Edits** **21.7%** Video **1 fav**
- [Staying Safe Online \(21/10/2010\) \[Video\]](#) All items | LSE Public lectures and events | Video **32 Edits** **0.4%** Video
- [BBC News Overnight 2011 with Gavin Grey](#) Media files in www.youtube.com **429 Edits** **74.0%** Video

38 channels 4310 episodes (66 have been corrected) 16175 corrected words (Edits) 933 searches 読者入力数 28

[Recently corrected episodes](#) - [Ranking](#) - [List of channels](#) - [Register contents](#)
Services: [Nico Nico Douga](#) | [YouTube](#) | [Ustream](#) | [Podcast](#)

What is PodCastle?

PodCastle is a service that enables searching of speech data such as podcasts, individual audio or video files on the web, and video clips on video sharing services (Nico Nico Douga, YouTube, and Ustream). Speech data are converted to text data by using an automatic speech

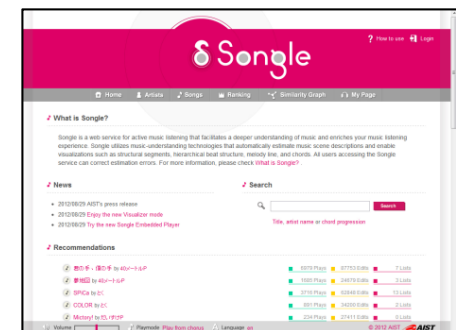
PodCastle (Web service for speech retrieval)

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



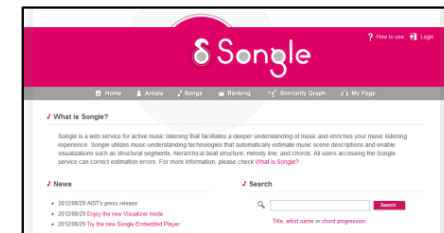


Our Goal

- ❑ 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**

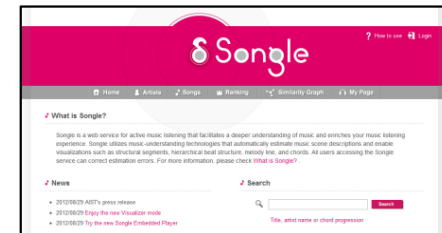
Originality

We collect and amplify voluntary contributions by anonymous users to **improve user experiences**

PodCastle for speech



Songle for music





PodCastle for speech



How to use - Login

searching podcasts
reading podcasts
annotating podcasts

Podcastle

Search [Advanced search](#)

Recommended Queries: [bit our information times called problem things talking](#)

Recommend

- [AIST - Paving the way to the Future](#) Media files in www.youtube.com 260 Edits 28.4 % Video 1 fav
- [CrestMuse: demo video \(4/5\)](#) Media files in www.youtube.com 142 Edits 21.7 % Video 1 fav
- [Staying Safe Online \(21/10/2010\) \[Video\]](#) All items | LSE Public lectures and events | Video 32 Edits 0.4 % Video
- [BBC News Overnight 2011 with Gavin Grey](#) Media files in www.youtube.com 429 Edits 74.0 % Video

38 channels 4310 episodes (66 have been corrected) 16175 corrected words (Edits) 933 searches 話者入力数 28

PodCastle

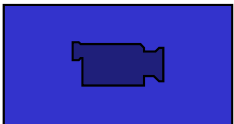


PodCastle (<http://podcastle.jp>)

- 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

Firefox

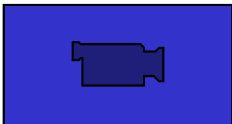
Podcastle www.youtube.com AIST - Paving the way to the Future AIST is one of the largest public research in [How to use](#) - [Login](#)

[History](#) [Delete](#) [Details](#) [Shortcut keys](#) [Map](#) [Comment](#) ver 1.0 241 Edits 27.6% Video 1 fav ☆

Full text Candidates Play Stop 01:07/16:27 Adjust the playback offset 27.6% Save

agency of industrial science and technology in the ministry
agency of industrials science an technology in the ministry
industrial signs of
and
a
in
uh

Corrected (Blue indicates that an error was corrected)
Confirmed (Green indicates that a speech recognition result was confirmed to be correct)
Degree of speech-recognition reliability (Red indicates that a speech recognition result might include an error)





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)

■ Corrected (Blue indicates that an error was corrected)
■ Confirmed (Green indicates that a speech recognition result was confirmed to be correct)
■ Degree of speech-recognition reliability (Red indicates that a speech recognition result might include an error)



Performance Improvement by Users

- ❑ Anonymous users can **find and correct** errors
- ❑ Novel effective approach for **improving ASR with end users** on errors
- ❑ Novel effective approach for **improving ASR with end users** on errors

[Ogata, Goto, 2004]

Corrected errors can be used for

- Improving **retrieval performances** by correct indices
- Improving **recognition performances** by machine learning (adaptation/training)

■ Corrected (Blue indicates that an error was corrected)
■ Confirmed (Green indicates that a speech recognition result was confirmed to be correct)
■ Degree of speech-recognition reliability (Red indicates that a speech recognition result might include an error)



History

❑ PodCastle Project (<http://en.podcastle.jp>)

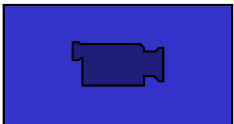
- 2006/01 Started the project
- 2006/12 Released to the public
The world's first speech retrieval using crowdsourcing
- 2008/06 Press release
Released to the public
Reported in TV news, newspapers, web news, etc.



History

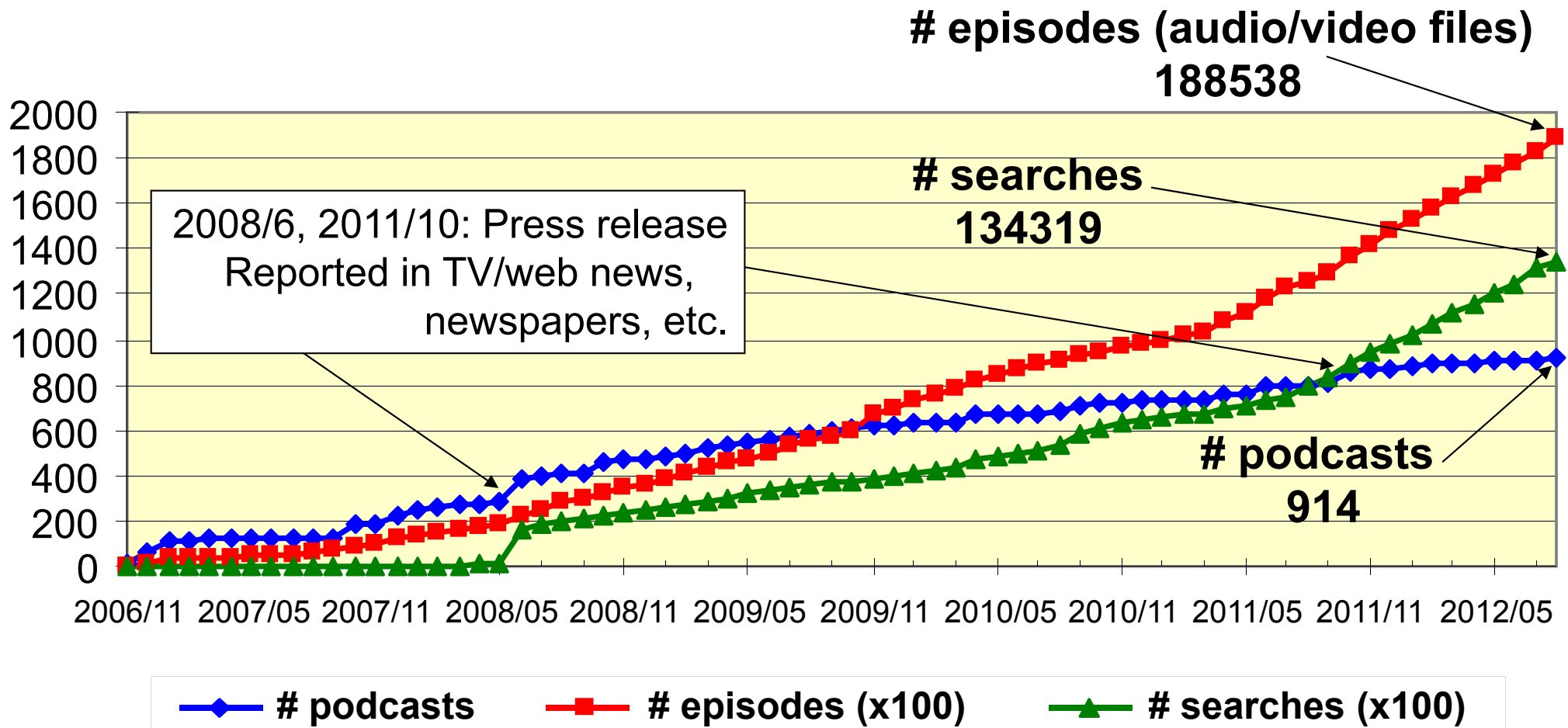
❑ PodCastle Project (<http://en.podcastle.jp>)

- 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](#)



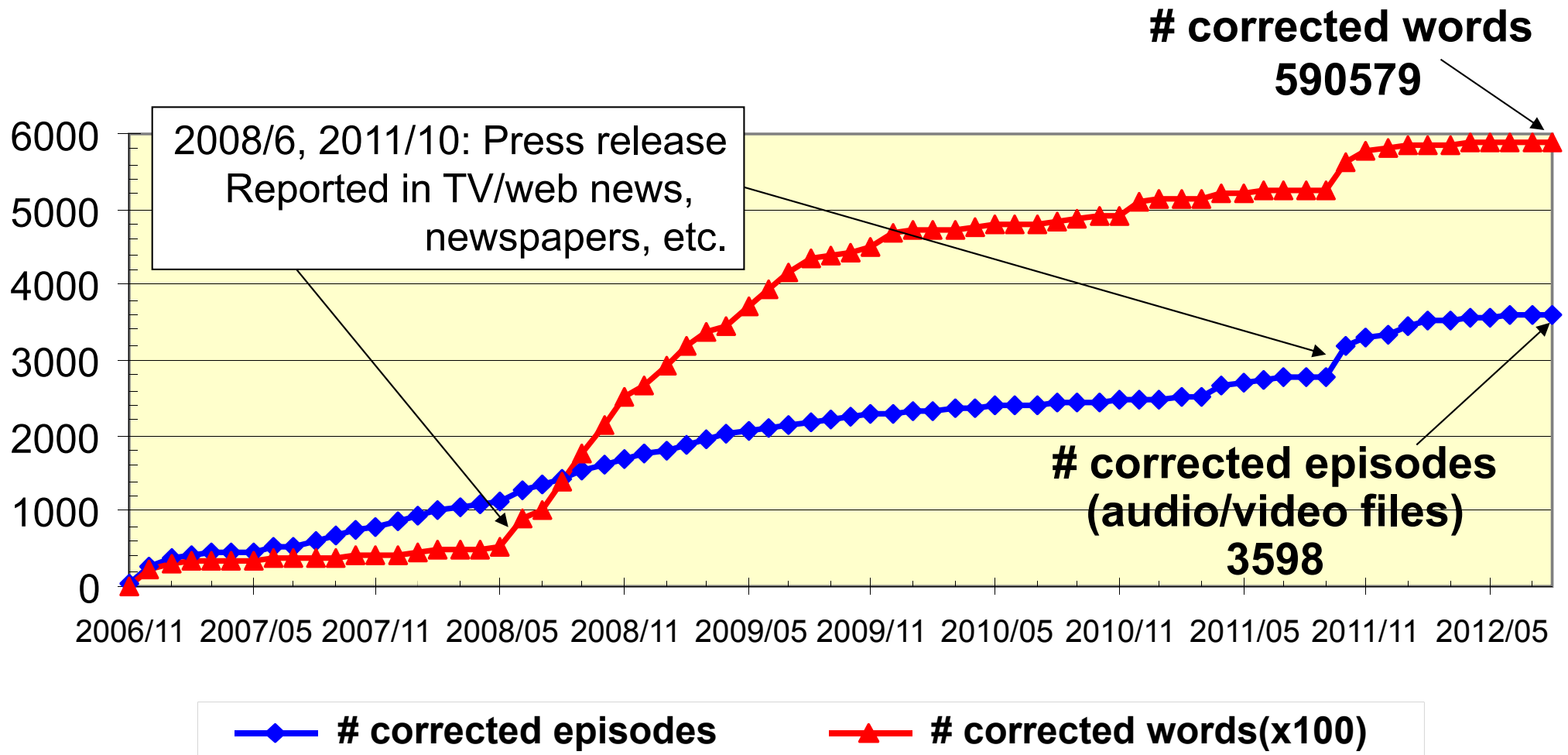
How Widely Used Is PodCastle?

Size of data stored in PodCastle



How Widely Used Is PodCastle?

Amount of corrections





How Widely Used Is PodCastle?

❑ As of September 1, 2012

- # registered channels 914
- # registered episodes (MP3s) 188538
- # corrected episodes (MP3s) 3598
- # corrected words 590579

❑ 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

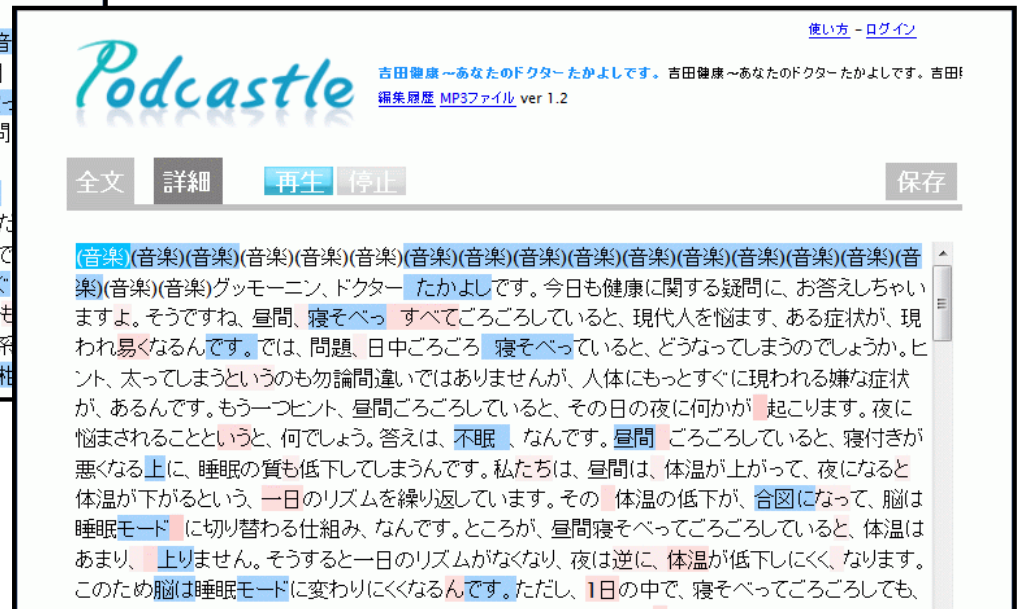


Podcastle 吉田健康～あなたのドクターたかよしです。吉田健康～あなたのドクターたかよしです。吉田健康～あなたのドクターたかよしです。編集履歴 MP3ファイル ver 1.01

全文 詳細 再生 停止 保存

(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)グッモーニング、ドクターたかよしです。今日も健康に関する疑問に、お答えし ちゃいますよー。そうですね。実は、朝一番にあるにおいを嗅ぐと、眠気 が吹っ飛んで、脳が活性化 します。では、いきなり問題。それは、何の香りでしょうか。今日は、三択 問題ですよ。一番、オレンジなどの、柑橘系の香り、二番、ラベンダーのあまーい 香り、三番、ジャスミンの エキゾチック な香り。いいですか。目が覚めて脳が活性化するのは、一番オレンジ、二番 ラベンダー、三番、ジャスミン、さあ、どれでしょうか。三番のジャスミンを選んだ方、大ヒットした、あの歌を思い出 してください。○四人GOでえー (は同十九分、えー。そうですね。ジャスミンは、脳の活動を穏やかにし、眠くする 作用が、あります。ですから、夜寝る前に嗅ぐのはお勧めですが、朝ジャスミンの香りを嗅い たら、二度寝してしまいます。二番のラベンダーも全く同じで、眠りを誘う作用があるので、朝には不向きです。ということで、え答えは、一番の柑橘系の香り、なんです。柑橘系の香りをかぐと、全身は どうなるのか。さあ今日のポイントですよ。柑

After learning
corrected errors



Podcastle 吉田健康～あなたのドクターたかよしです。吉田健康～あなたのドクターたかよしです。吉田健康～あなたのドクターたかよしです。編集履歴 MP3ファイル ver 1.2

全文 詳細 再生 停止 保存

(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)(音楽)グッモーニング、ドクター たかよしです。今日も健康に関する疑問に、お答えし ちゃいますよ。そうですね。昼間、寝そべっ すべてごろごろしていると、現代人を悩ます、ある症状が、現われ 易くなるんです。では、問題、日中ごろごろ 寝そべっていると、どうなるのでしょうか。ヒント、太ってしまうというのも勿論間違いではありませんが、人体にもっとすぐに現われる嫌な症状が、あるんです。もう一つヒント、昼間ごろごろしていると、その日の夜に何かか 起こります。夜に悩まされることという、何でしょう。答えは、不眠、なんです。昼間 ごろごろしていると、寝付きが悪くなる上 に、睡眠の質も低下してしま います。私たちは、昼間は、体温が上がって、夜になると体温が下がるという、一日のリズムを繰り返しています。その 体温の低下が、合図 になって、脳は睡眠モード に切り替わる仕組み、なんです。ところが、昼間寝そべってごろごろしていると、体温はあまり、 上がりません。そうすると一日のリズムがなくなり、夜は逆に、体温が低下しにく くなります。このため脳は睡眠モードに変わりにくくなるんです。ただし、1日の中で、寝そべってごろごろしても、

Errors (blue words)
were reduced



Why Do Users Correct Errors?

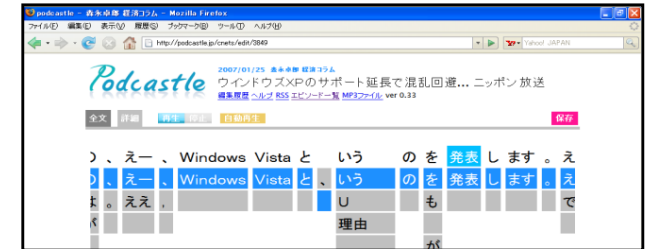
❑ Inferred motivations

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

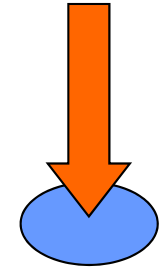
Discussion

□ Framework for **amplifying user contributions**

- Typical Web 2.0 services

Improvements are limited to

an item directly contributed by users



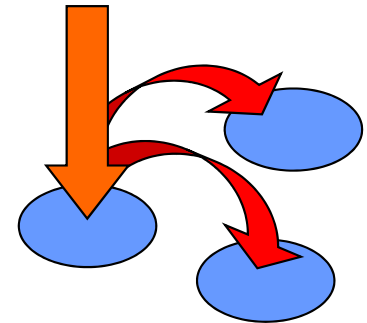
- **Beyond Web 2.0:** Advantage of PodCastle

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**

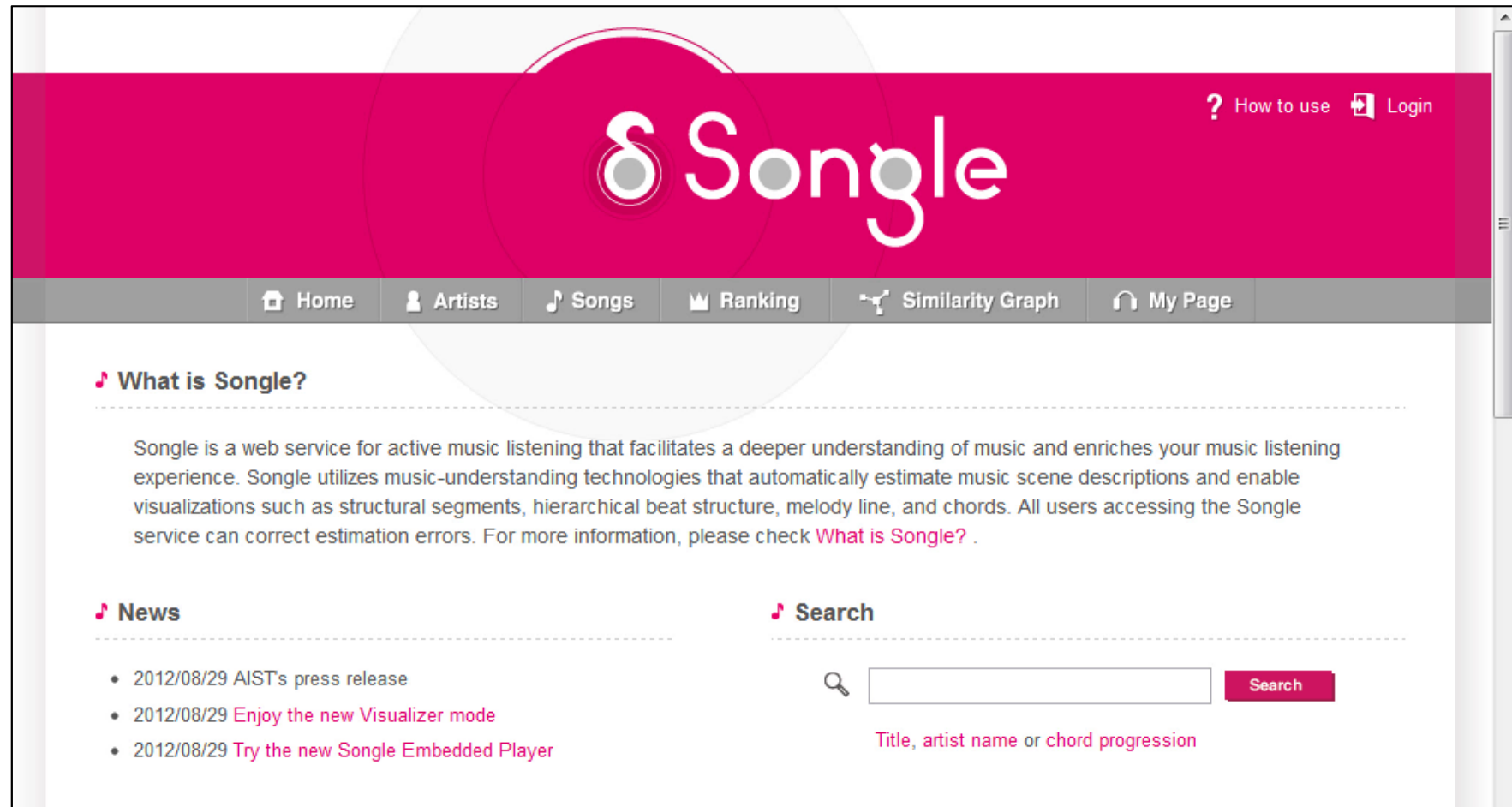
- because **our ASR system** is getting better and better

❑ **We would eventually obtain**

the best ASR system in the world



Songle for music



The screenshot shows the Songle website interface. At the top, there is a red header with the Songle logo (a stylized 'S' with a musical note) and the text 'Songle'. To the right of the logo are links for '? How to use' and 'Login'. Below the header is a grey navigation bar with icons and labels for 'Home', 'Artists', 'Songs', 'Ranking', 'Similarity Graph', and 'My Page'. The main content area is white and features a section titled 'What is Songle?' with a dashed line separator. Below this is a paragraph of text describing the service. To the right of this text is a 'Search' section with a magnifying glass icon, a search input field, and a red 'Search' button. Below the search input field is the text 'Title, artist name or chord progression'. On the left side of the main content area, there is a 'News' section with a dashed line separator and a list of three news items, each starting with a date and a link.

What is Songle?

Songle is a web service for active music listening that facilitates a deeper understanding of music and enriches your music listening experience. Songle utilizes music-understanding technologies that automatically estimate music scene descriptions and enable visualizations such as structural segments, hierarchical beat structure, melody line, and chords. All users accessing the Songle service can correct estimation errors. For more information, please check [What is Songle?](#)

News

- 2012/08/29 [AIST's press release](#)
- 2012/08/29 [Enjoy the new Visualizer mode](#)
- 2012/08/29 [Try the new Songle Embedded Player](#)

Search

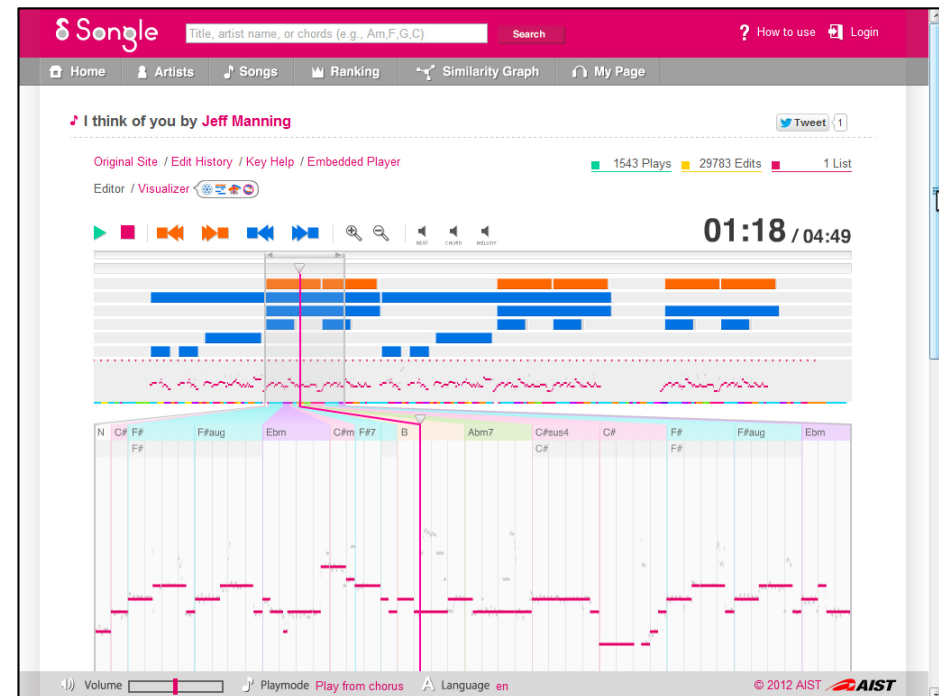
[Search](#)

Title, artist name or chord progression

Songle (http://songle.jp)

❑ Web service for **active music listening**

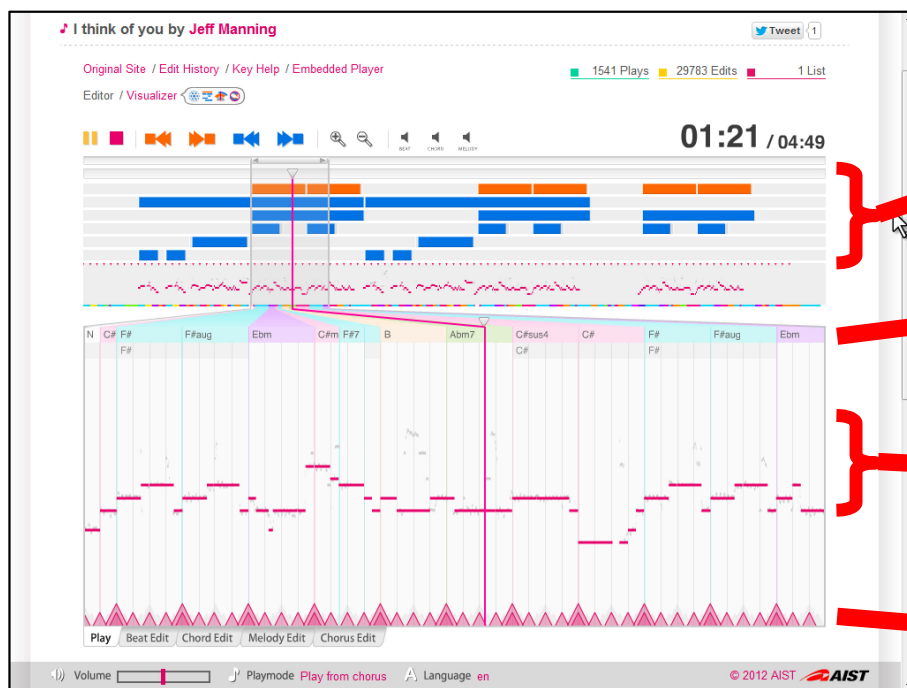
- Allow **anonymous users** to enjoy **any songs (MP3 files) available on the web** by using **active music listening interfaces**



Songle (<http://songle.jp>)

❑ Use automatic music-understanding technologies

- Estimate four major types of music scene descriptions

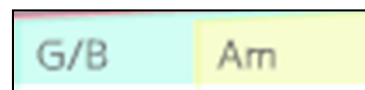


Music structure

(chorus / repeated sections)

Chords

(root note and chord type)



Melody line

(F0 of the vocal melody)

Beat structure

(musical beats and bar lines)

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



Demonstration (Visualization)

♪ I think of you by Jeff Manning Tweet 1

[Original Site](#) / [Edit History](#) / [Key Help](#) / [Embedded Player](#) 1541 Plays 29783 Edits 1 List

Editor / Visualizer

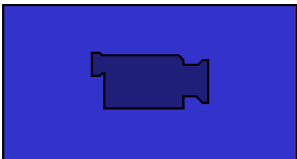
01:21 / 04:49

The visualization interface displays a piano roll with notes in blue and orange, a waveform in red, and a chord progression below. The chord progression is as follows:

Chord	Chord	Chord	Chord	Chord	Chord	Chord	Chord	Chord	Chord	Chord	Chord	Chord	
N	C#	F#	F#aug	Ebm	C#m	F#7	B	Abm7	C#sus4	C#	F#	F#aug	Ebm

At the bottom, there are tabs for **Play**, **Beat Edit**, **Chord Edit**, **Melody Edit**, and **Chorus Edit**.

Volume Playmode **Play from chorus** Language **en** © 2012 AIST



Demonstration (Visualization)

♪ I think of you by Jeff Manning Tweet 1

[Original Site](#) / [Edit History](#) / [Key Help](#) / [Embedded Player](#) 1541 Plays 29783 Edits 1 List

Editor / Visualizer 🌐 🎛️ 🎧

01:21 / 04:49

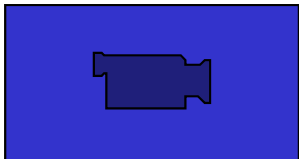
BEAT CHORD MELODY

Songle facilitates deeper understanding of music
by visualizing music scene descriptions

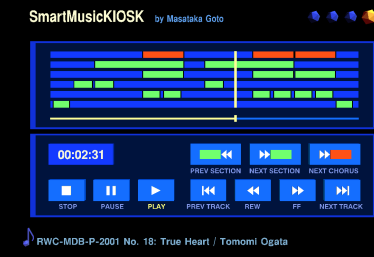
Play Beat Edit Chord Edit Melody Edit Chorus

Volume 🔊 Playmode [Play from chorus](#) Language [en](#)

© 2012 AIST



Songle with SmartMusicKIOSK

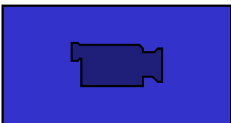


❑ Implement all functions of SmartMusicKIOSK

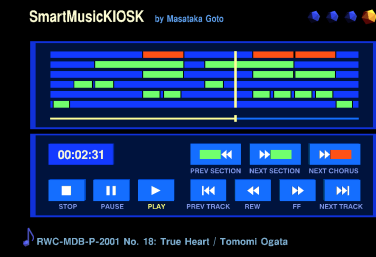
- Jump and listen to the **chorus** with just a push of a button

[Goto, ACM UIST 2003]

A screenshot of the SmartMusicKIOSK interface for the song "I think of you by Jeff Manning". The interface shows a music player with a progress bar at the top, a time display showing 01:21 / 04:49, and various playback controls. Below the controls, there are buttons for 'PREV SECTION', 'NEXT SECTION', and 'NEXT CHORUS'. The interface is titled 'I think of you by Jeff Manning'. The main area displays a music score with a timeline and a chord progression. The chorus is highlighted in red and labeled 'Chorus' in red text. The repeated sections are highlighted in blue and labeled 'Repeated sections' in blue text. The chord progression at the bottom is: N C# F# F#aug Ebm C#m F#7 B Abm7 C#sus4 C# F# F#aug Ebm.



Songle with SmartMusicKIOSK

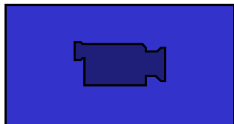


❑ Implement all functions of SmartMusicKIOSK

- Jump and listen to the **chorus** with just a push of a button

[Goto, ACM UIST 2003]

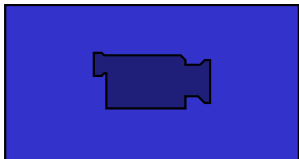
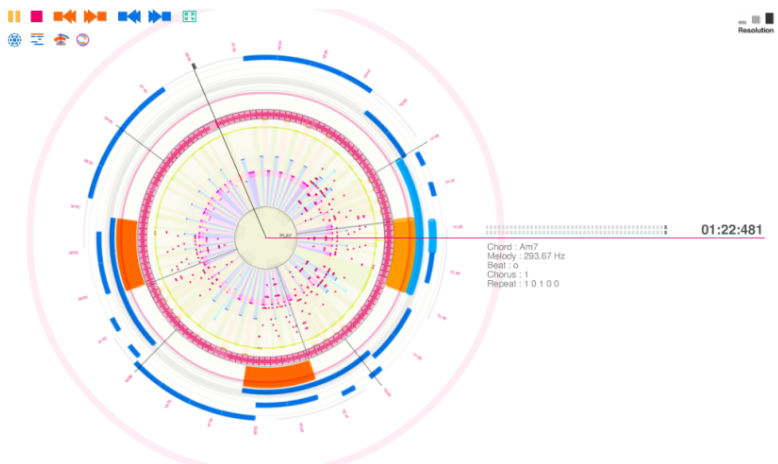
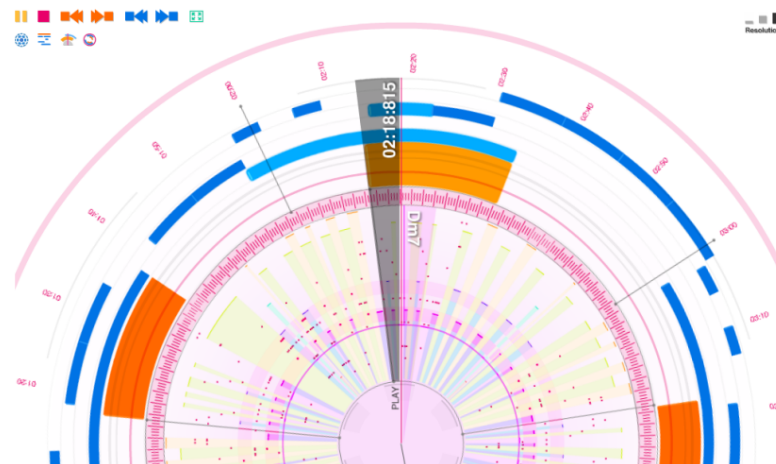
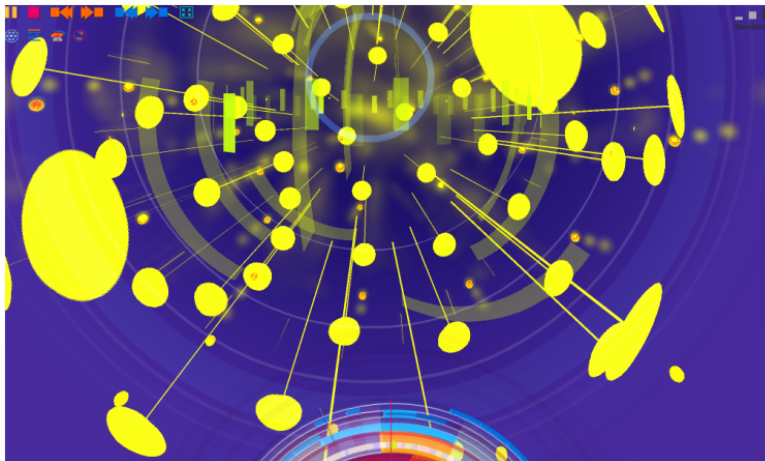
A screenshot of the Songle interface for the song 'I think of you by Jeff Manning'. The interface shows a timeline with repeated sections highlighted in blue. A red box highlights the text: 'Songle makes it easier for a user to find desired parts of a piece'. Below the timeline, the text 'Repeated sections' is written in blue. At the bottom, there is a chord progression: N C# F# F#aug Ebm C#m F#7 B Abm7 C#sus4 C# F# F#aug Ebm. A mouse cursor is visible on the right side of the interface.



New Function: Visualizer

Visualizer Mode

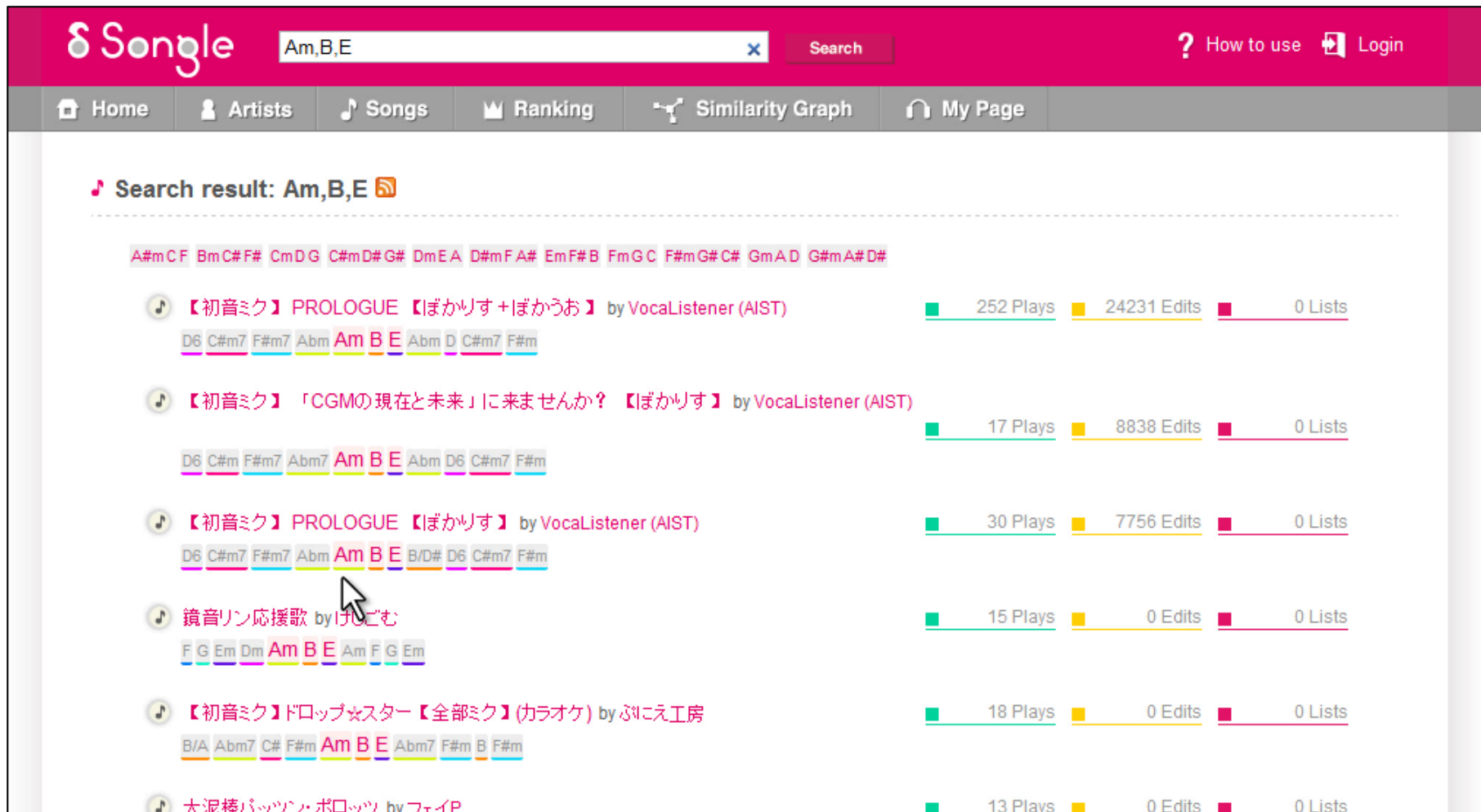
- Four types of animation rigidly synchronized with music



New Function: Content-based Retrieval

❑ Chord Progression Search

- Listen to songs having the same sequence of **chord names**



The screenshot shows the Songle website interface. At the top, there is a search bar with the text "Am,B,E" and a "Search" button. To the right of the search bar are links for "? How to use" and "Login". Below the search bar is a navigation menu with icons and labels for "Home", "Artists", "Songs", "Ranking", "Similarity Graph", and "My Page".

The main content area displays the search results for "Am,B,E". At the top of the results, there is a list of chord names: A#mCF BmC#F# CmDG C#mD#G# DmEA D#mFA# EmF#B FmGC F#mG#C# GmAD G#mA#D#.

The search results are listed as follows:

Song Title	Artist	Plays	Edits	Lists
【初音ミク】 PROLOGUE 【ばかりす+ばかりお】	VocaListener (AIST)	252	24231	0
【初音ミク】 「CGMの現在と未来」に來ませんか? 【ばかりす】	VocaListener (AIST)	17	8838	0
【初音ミク】 PROLOGUE 【ばかりす】	VocaListener (AIST)	30	7756	0
鏡音リン応援歌	けいごむ	15	0	0
【初音ミク】ドロップ☆スター 【全部ミク】(カラオケ)	ぶにえ工房	18	0	0
大泥棒パッツン・ボロツツ	フェイP	13	0	0

Each result includes a chord progression snippet below the title. For example, the first result shows "D6 C#m7 F#m7 Abm Am B E Abm D C#m7 F#m". A mouse cursor is visible over the "Am B E" part of the third result's chord progression.

New Function: Songle Embedded Player

❑ Songle Embedded Player

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



[Japanese](#)

Songle Embedded Player

Demonstration of Music-Synchronized Animation

Let's push the play button below!

A web service for active music listening "Songle" (<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-synchronized animation** in the background of your web page. This page is intended to demonstrate this function.

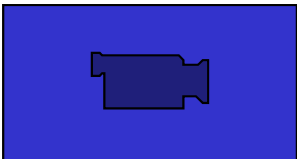
Note that the music-synchronized animation cannot be shown on Internet Explorer 8 or below because HTML5 is not supported. [Chrome browser](#) is recommended.

This function became available after [AIST's press release](#) on August 29th, 2012.

Examples of web pages demonstrating the Songle Embedded Player:

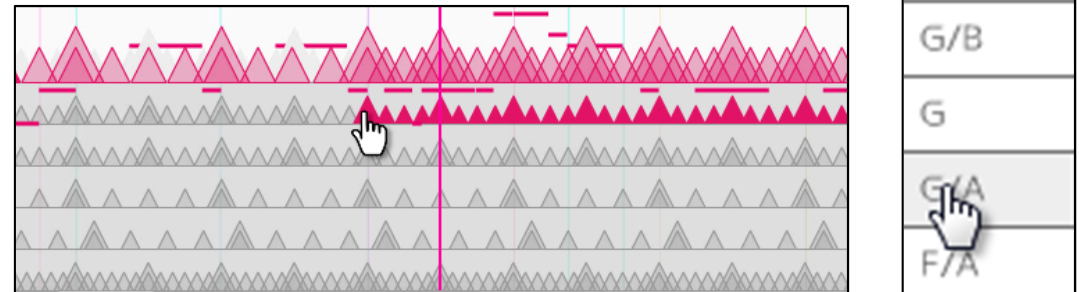
- [Miku 5th Anniversary](#)
- [PIAPRO blog](#)

How to embed the Songle Embedded Player:



Performance Improvement by Users

- ❑ 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



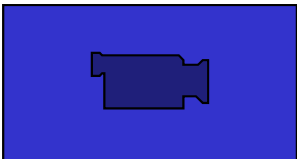
by providing an alternative description



Demonstration (Error Correction)

Music structure

The screenshot displays a music production software interface with a piano roll and chord editor. The top right corner shows a time display of 01:14 / 04:49. The interface includes a toolbar with play, stop, and navigation buttons, as well as volume and solo controls for BEAT, CHORD, and MELODY. The piano roll shows a melody line with a red cursor and a chord editor below it. The chord editor displays a sequence of chords: Ebm, Bbm, B, C#, B, C#, B, B, N, F#, Ebm, B, C#, F#, Ebm, B, F#, F#, Ebm, F#. The bottom of the interface features a navigation bar with buttons for Play, Beat Edit, Chord Edit, Melody Edit, and Chorus Edit, along with Undo, Redo, Insert, Delete, and Save buttons.

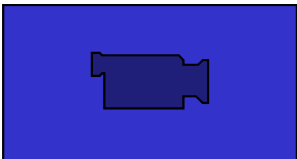




Demonstration (Error Correction)

Beat structure

The screenshot displays a music software interface with a top toolbar containing play, stop, and navigation icons, along with a timer showing 01:14 / 04:49. The main area is divided into several sections: a piano roll at the top with blue and orange notes, a waveform below it, and a chord progression section at the bottom. The chord progression is labeled with various chords: N, C#, F#, F#aug, Ebm, C#m, F#7, F#7, B, Abm7, C#sus4, and C#. Below the chord labels is a series of red triangles representing the beat structure. At the bottom, there are buttons for 'Play', 'Beat Edit', 'Chord Edit', 'Melody Edit', 'Chorus Edit', 'Undo', 'Redo', 'Half-beat shift', 'Smoothing', and 'Save'. A 'Tapping mode' checkbox is also visible on the right side of the chord progression area.

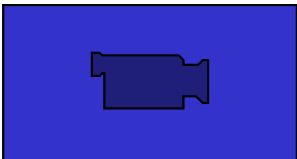




Demonstration (Error Correction)

□ Melody line

The screenshot displays a music software interface with a timeline at the top showing playback controls and a time display of 01:14 / 04:49. Below the timeline, there are several tracks: a piano track with a red waveform, a MIDI track with a pink piano roll, and a chord track with a grid of colored chords. The chord track shows a progression: C#, F#, F#aug, Ebm, C#m, F#7, F#7, B, Abm7, C#sus4, and C#. The piano roll shows a melody line with pink notes. The interface includes a bottom menu with options: Play, Beat Edit, Chord Edit, Melody Edit, Chorus Edit, Undo, Redo, Copy, Paste, and Save. There are also checkboxes for 'Detailed mode' and 'Snap to beat'.

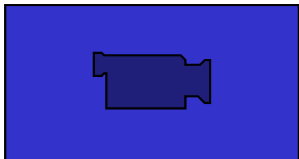




Demonstration (Error Correction)

Chords

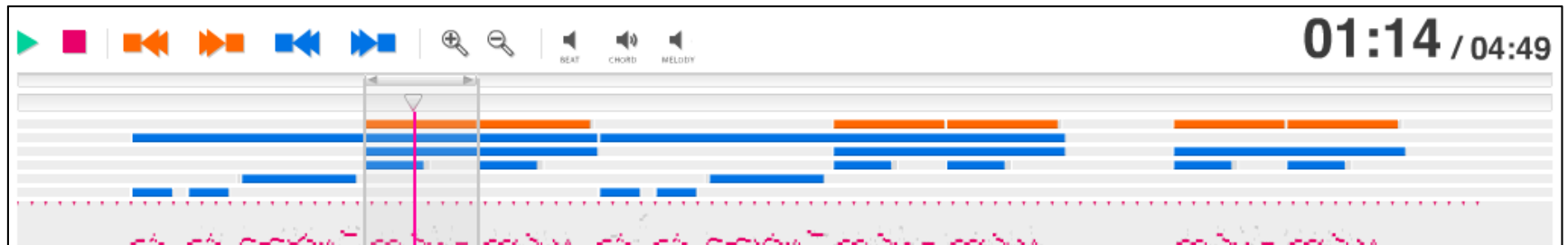
The screenshot shows a music software interface with a piano roll at the top and a chord list at the bottom. The piano roll displays a sequence of chords: N, C#, F#, F#aug, Ebm, C#m, F#7, F#7, B, Abm7, C#sus4, and C#. The chord list at the bottom shows the following options: C#m, C#m7, F#/C#, C#6, C#, F#7 (highlighted), F#, C#dim, Bbdim7, A/C#, N, and Other. The interface also includes a toolbar with playback controls, a timer showing 01:14 / 04:49, and a bottom menu with options like Play, Beat Edit, Chord Edit, Melody Edit, Chorus Edit, Undo, Redo, Insert, Delete, and Save.



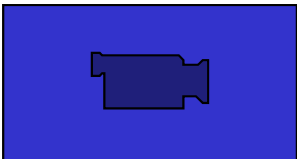
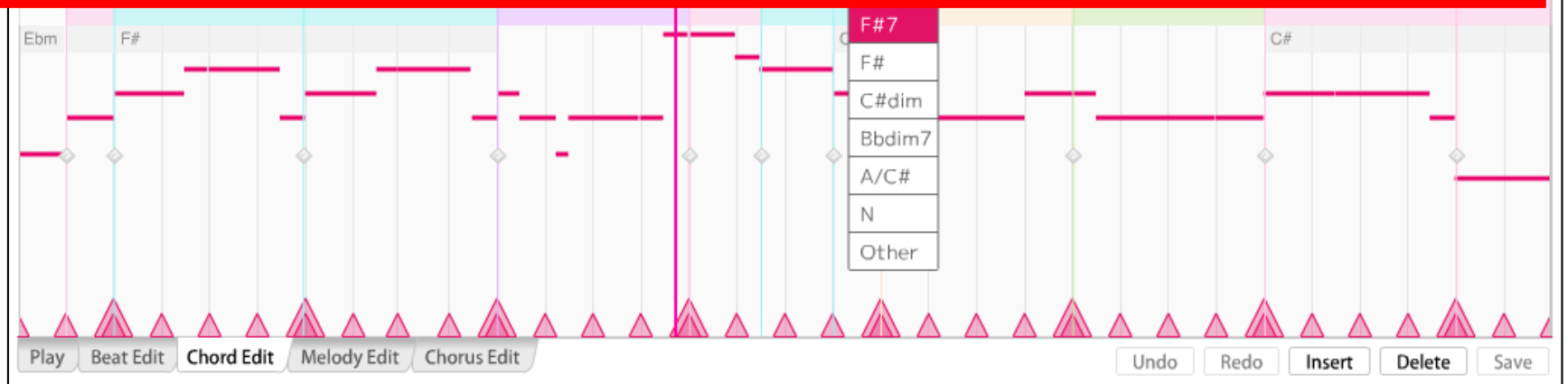


Demonstration (Error Correction)

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

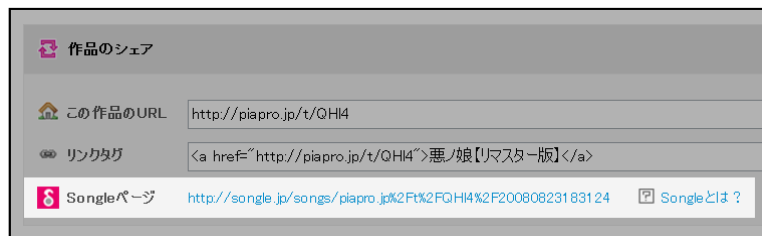
❑ Collaboration with **PIAPRO** (CGM content site)

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



Chorus sections



Conclusion

PodCastle for speech

The screenshot shows the PodCastle website. At the top, it says "searching podcasts reading podcasts annotating podcasts" next to the PodCastle logo. Below the logo is a search bar with a "Search" button and a link to "Advanced search". Underneath, there are "Recommended Queries" such as "hit our information times called problem things talking". A "Recommend" section lists several items with their respective edit counts and video links. At the bottom, there are statistics: "38 channels", "4310 episodes (66 have been corrected)", "16175 corrected words (68%)", and "933 searches". A footer section titled "What is PodCastle?" explains that it is a service for searching speech data from various sources.

Songle for music

The screenshot shows the Songle website. The header is pink with the Songle logo and navigation links for Home, Artists, Songs, Ranking, Similarity Graph, and My Page. Below the header, there is a "What is Songle?" section explaining that it is a web service for active music listening that provides deeper understanding and visualization of music. A "News" section lists recent updates. A "Search" box is available with a search button and a prompt to search by "Title, artist name or chord progression". A "Recommendations" section lists several songs with their play counts, edit counts, and list counts. At the bottom, there is a volume control slider, a playmode selector set to "Play from chorus", and a language selector.



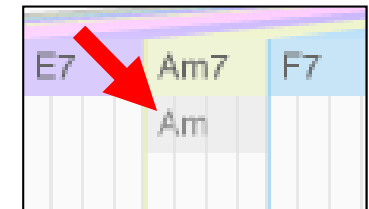
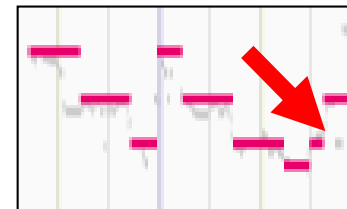
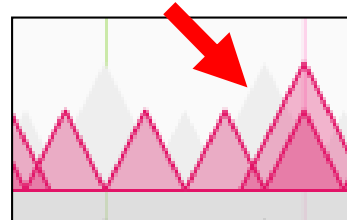
Social Contributions

- ❑ **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

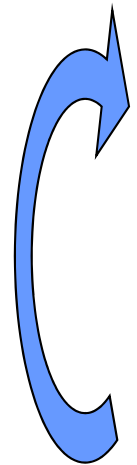
Academic Contributions

PodCastle and Songle

- New research approach to speech recognition and music understanding

Better user experiences

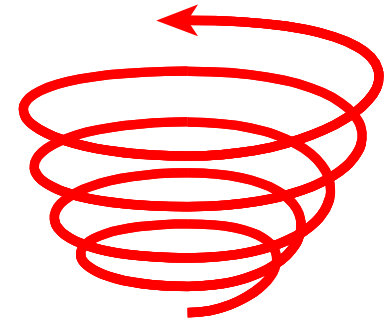
- Designed to set into motion a *positive spiral*



(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)



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

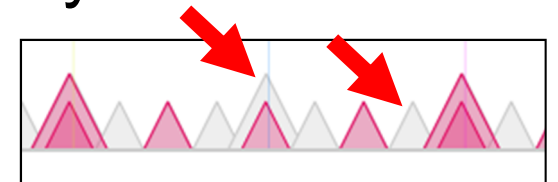
[Turnbull *et al.*, '07] [Mandel *et al.*, '07] [Law *et al.*, '07][Lee, '10][Mandel *et al.*, '10]

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

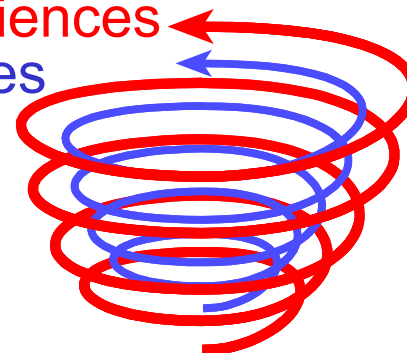
flu	shot	locator	
	shot	will	cater
low	lusaka	locator	
flu	longshot	while	



Future work

- Use **corrected errors** and **machine learning techniques** to improve **music-understanding technologies** for Songle

Better user experiences
Better technologies





Let's Enjoy!



<http://en.podcastle.jp>



<http://songle.jp>

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 Ito 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/>