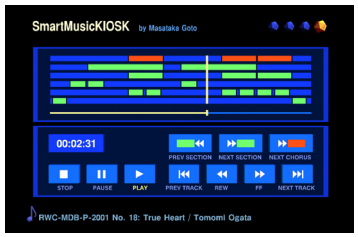
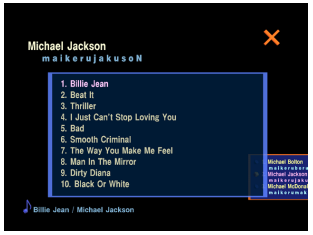
Toward Music Listening Interfaces in the Future

AIST (National Institute of Advanced Industrial Science and Technology)

Masataka Goto





2010/10/19 Microsoft Research Asia Faculty Summit 2010

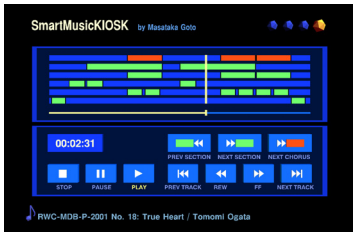
Our Goal

- Enrich end-users' music listening experiences
  by using music understanding,
  speech interaction, and
  humanoid robot technologies

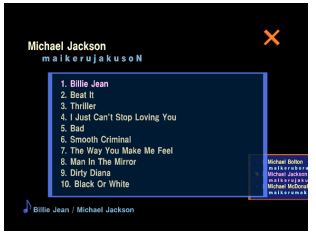
- Change music listening into
  a more active, immersive experience

Music Listening Interfaces in the Future

- **Natural user interaction for music can be enriched by**
  - Music understanding technology
    - Content-based analysis/visualization



  - Speech interaction technology
    - Nonverbal interaction with speech recognition



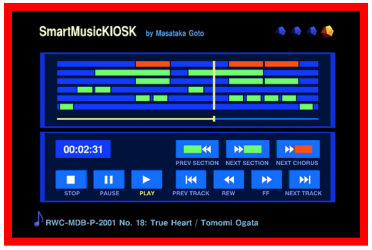
  - Humanoid robot technology
    - Rigidly-synchronous character

Music Listening Interfaces in the Future

- **Natural user interaction for music can be enriched by**

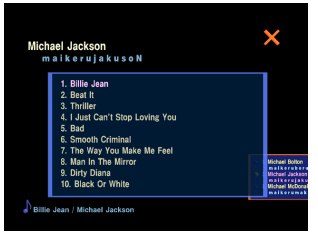
  - Music understanding technology

    Content-based analysis/visualization



  - Speech interaction technology

    Nonverbal interaction with speech recognition



  - Humanoid robot technology

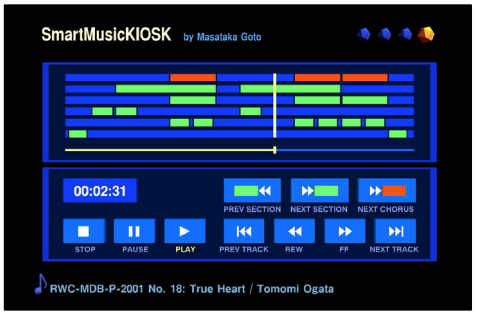
    Rigidly-synchronous character

Our Research Approach

Active Music Listening Interfaces

- Building Active Music Listening Interfaces
  that enable *non-musician users*
  to enjoy music in more active ways

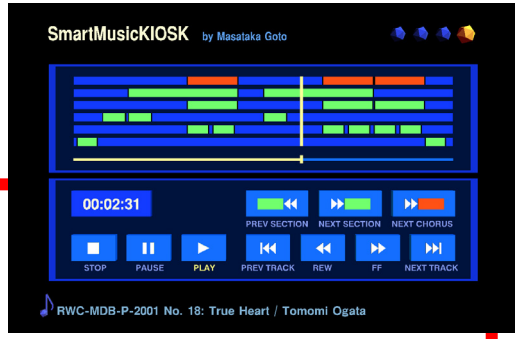
- Two interfaces
  - **SmartMusicKIOSK**
  - **LyricSynchronizer**



SmartMusicKIOSK [Goto, 2002-2006]

- **One of the easiest active interaction**
  - Skip musical pieces of no interest
    by pressing the “NEXT TRACK” button

- **More advanced active interaction?**
  - Skip sections of no interest within a song

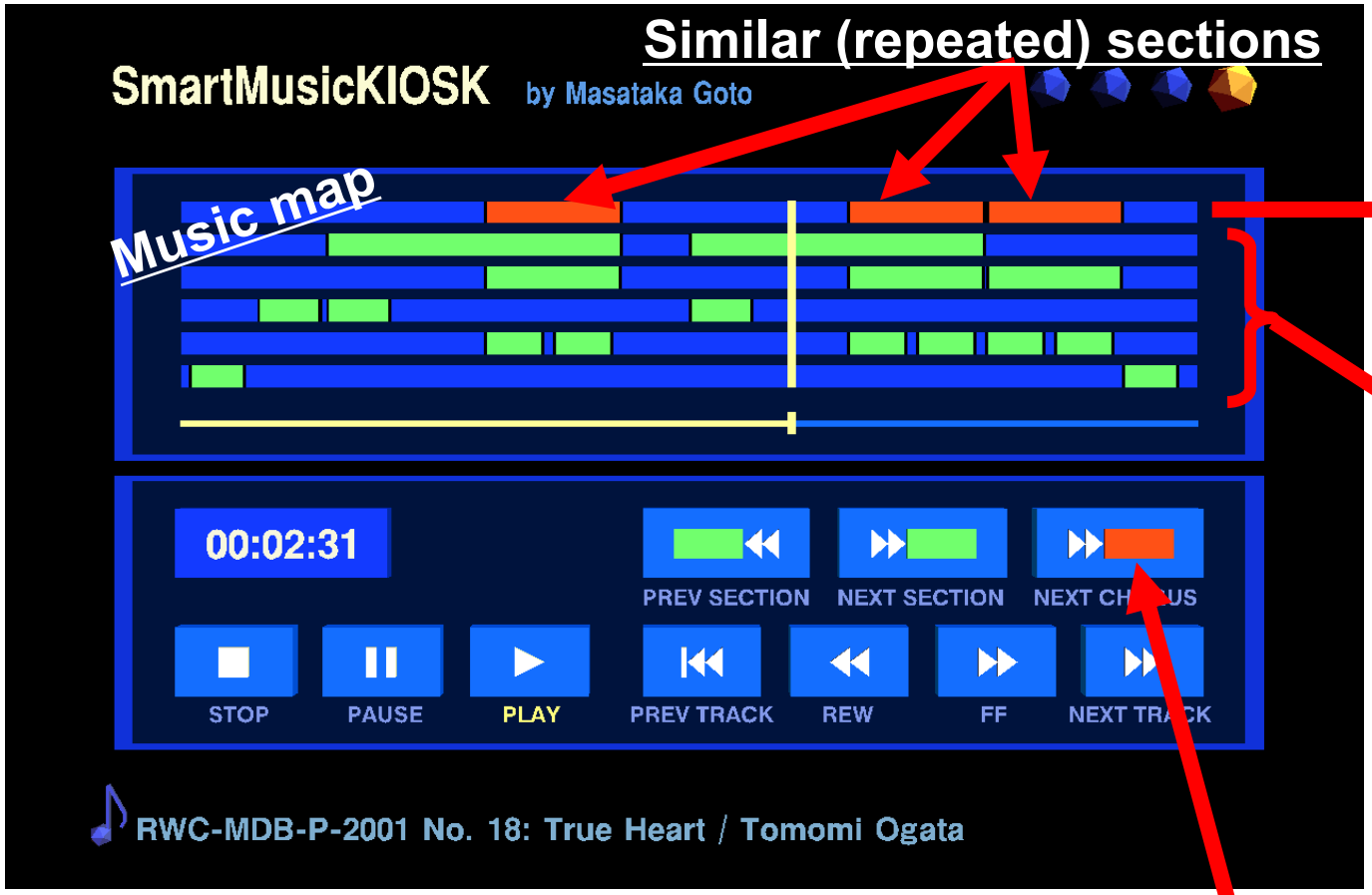


INTERFACE:
SmartMusicKIOSK:
Music listening station with a chorus-search function

TECHNOLOGY:
Automatic chorus-section detection method

INTERACTION:
Change playback position while viewing “music map”

SmartMusicKIOSK [Goto, 2002-2006]



Chorus sections

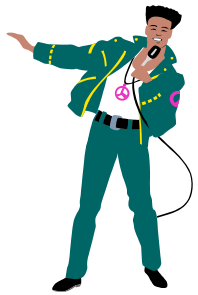
Repeated sections

"Jump to chorus" button

LyricSynchronizer

[Fujihara, Goto, Okuno, 2006-]

- **Reading/singing lyrics during music playback**
  - Refer to printed/displayed lyrics
  - Should keep track of the current playback position



- **More advanced active interaction?**
  - See/click the lyrics with the phrase being sung highlighted

INTERFACE:

LyricSynchronizer:

Synchronization of lyrics with music

TECHNOLOGY:

Automatic vocal extraction & synchronization method

INTERACTION:

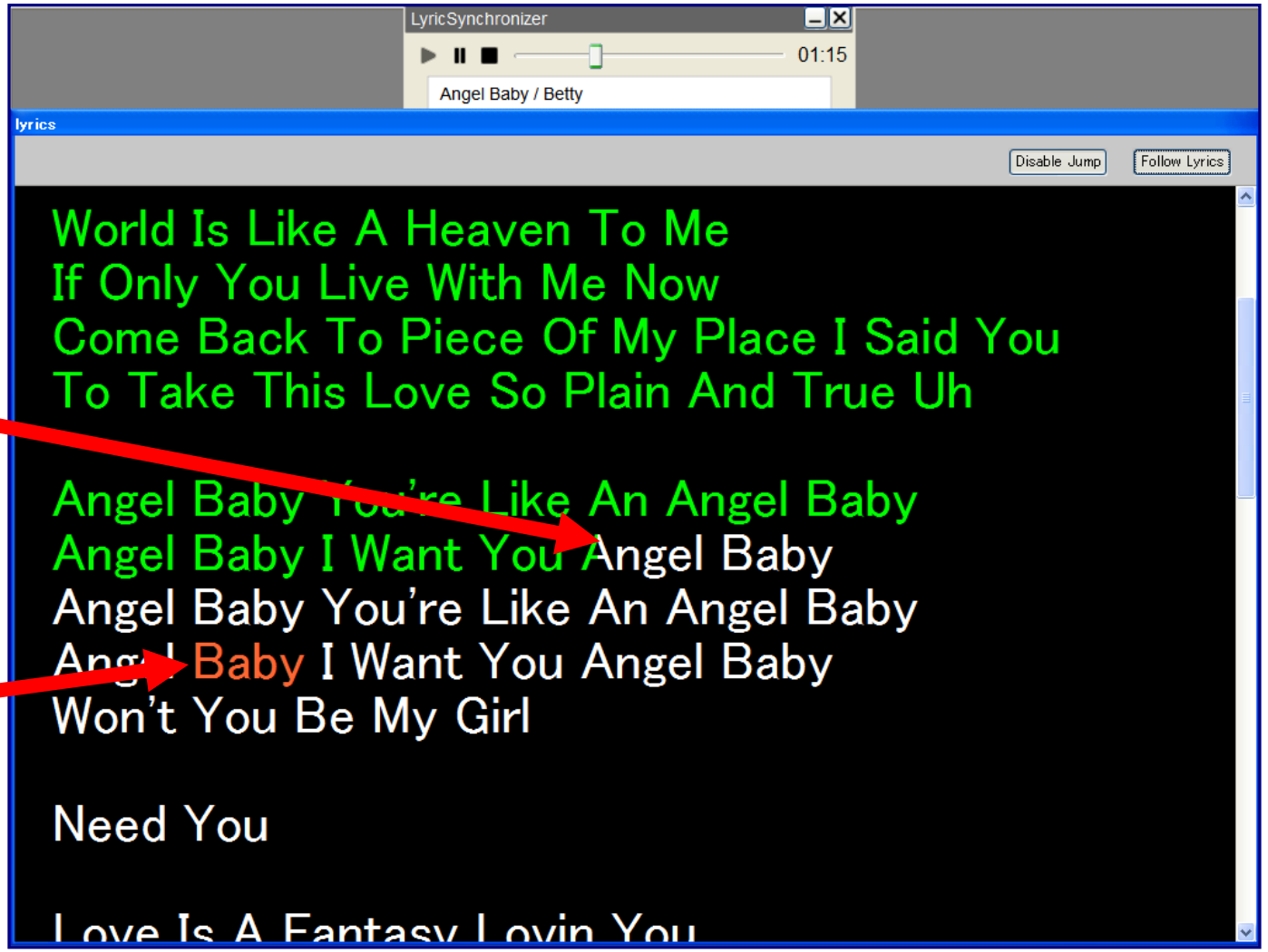
Click on a word in the lyrics to listen from that word

LyricSynchronizer

[Fujihara, Goto, Okuno, 2006-]

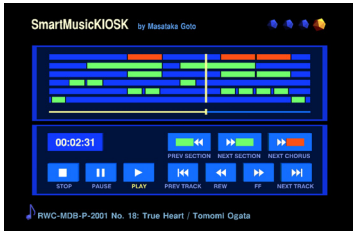
The current playback position

You can listen from a clicked word



Music Listening Interfaces in the Future

- **Natural user interaction for music can be enriched by**
  - Music understanding technology
    - Content-based analysis/visualization



  - Speech interaction technology
    - Nonverbal interaction with speech recognition

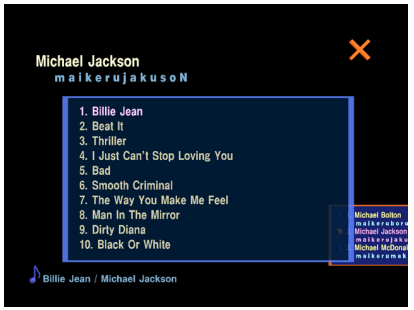
  - Humanoid robot technology
    - Rigidly-synchronous character

Our Research Approach

Speech Recognition Interfaces

- Building hands-free music listening interfaces
  that enable users
  to find and play back a musical piece

- Two interfaces
  - **Speech Completion**
  - **Speech Spotter**



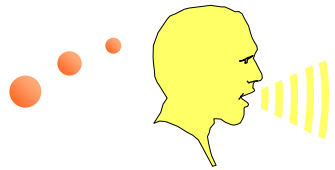
Speech Completion

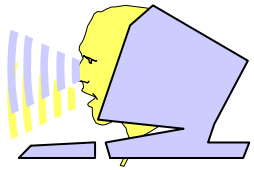
[Goto, Itou, Hayamizu, 2000-2004]

What is Speech Completion?

- Help a user enter an uncertain piece/artist name
  by completing the missing part
  of a partially uttered fragment

"Michael—" (Michael, uh…)





"Michael Jackson?"

Speech Completion

[Goto, Itou, Hayamizu, 2000-2004]

❑ Video Demonstration of Speech Completion

- Enter the *Japanese* names of musicians and songs

"Michael Jackson"

⇩

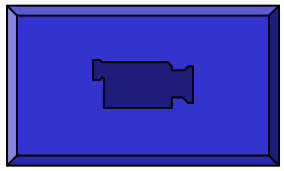
"MAIKERU JAKUSON"

(in Japanese)

"Michael—"

⇩

"MAIKERU—"

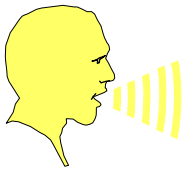


Speech Spotter

[Goto, Kitayama, Itou, Kobayashi, 2000-2004]

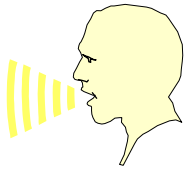
What is Speech Spotter?

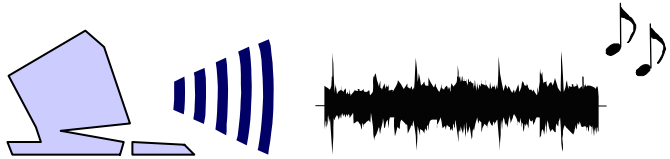
- Regard a user utterance as a command utterance
  only when it is intentionally uttered
  with a high pitch just after a filled pause (e.g., "er…")
  (prolonged vowel)



"Shall we listen to the song `Black or While' ?"

"Yeah! Uhm…, Black or White."





Speech Spotter

[Goto, Kitayama, Itou, Kobayashi, 2000-2004]

❑ Video Demonstration of Speech Spotter

- Enter voice commands for music-playback control

Speech Spotter

[Goto, Kitayama, Itou, Kobayashi, 2000-2004]

What is Speech Spotter?

- Regard a user utterance as a command utterance
  only when it is intentionally uttered
  with a high pitch just after a filled pause (e.g., "er…")
  (prolonged vowel)

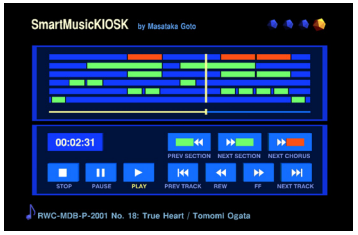
This combination is quite unnatural

= This does not appear in natural conversation

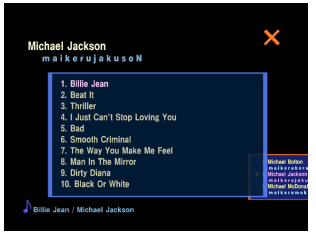
The system can easily find
this specially-designed unnatural utterance only

Music Listening Interfaces in the Future

- **Natural user interaction for music can be enriched by**
  - Music understanding technology
    - Content-based analysis/visualization



  - Speech interaction technology
    - Nonverbal interaction with speech recognition



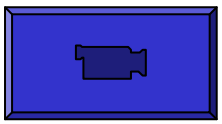
  - Humanoid robot technology
    - Rigidly-synchronous character

Our Research Approach

Humanoid Robot Interfaces

- Building immersive music listening interfaces
  that enable users
  to listen to a song while seeing a robot singer

- One example
  - **HRP-4C + VocaListener**
    + VocaWatcher



PROLOGUE 2010

[Kajita, Nakano, Goto, *et al.* 2010-]

HRP-4C + VocaListener + VocaWatcher

- **Two technologies to generate a natural singing voice and facial expressions by imitating a human singer**
  - VocaListener

    Technology to imitate the pitch and power of a human voice
  - VocaWatcher

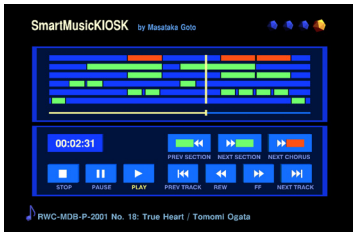
    Technology to imitate facial expressions of a human face

Music Listening Interfaces in the Future

- **Natural user interaction for music can be enriched by**

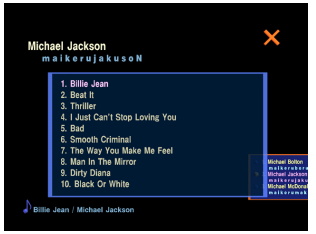
  - Music understanding technology

    Content-based analysis/visualization



  - Speech interaction technology

    Nonverbal interaction with speech recognition



  - Humanoid robot technology

    Rigidly-synchronous character

Conclusion

Summary

- Natural user interaction can be enriched by

  Content-understanding technology

  Content-based analysis/visualization

  Speech interaction technology

  Nonverbal interaction

  Humanoid robot technology

  Rigidly-synchronous character

  Web interaction technology

  User contributions

  Panel Discussion

Thank You

- **References** **(available at http://staff.aist.go.jp/m.goto/publications.html)**

M. Goto: *SmartMusicKIOSK: Music Listening Station with Chorus-Search Function*, ACM UIST 2003.

M. Goto: *A Chorus-Section Detection Method for Musical Audio Signals and Its Application to a Music Listening Station*, IEEE TASLP, 14(5), 1783-1794, 2006.

M. Goto: *Active Music Listening Interfaces Based on Signal Processing*, IEEE ICASSP 2007. (Invited Paper)

H. Fujihara, M. Goto, *et al.*: *Automatic Synchronization between Lyrics and Music CD Recordings Based on Viterbi Alignment of Segregated …*, IEEE ISM 2006.

M. Goto, K. Itou, K. Kitayama, and T. Kobayashi: *Speech-Recognition Interfaces for Music Information Retrieval: ``Speech Completion'' and ``Speech Spotter''*, ISMIR 2004.

M. Goto, K. Itou, and S. Hayamizu: *Speech Completion: On-demand Completion Assistance Using Filled Pauses for Speech Input Interfaces*, ICSLP 2002.

M. Goto, K. Kitayama, K. Itou, and T. Kobayashi: *Speech Spotter: On-demand Speech Recognition in Human-Human Conversation …*, ICSLP 2004.

M. Goto, K. Itou, and T. Kobayashi: *Speech Interface Exploiting Intentionally-Controlled Nonverbal Speech Information*, ACM UIST 2005.

Acknowledgments

- **Hiromasa Fujihara** (for LyricSynchronizer)
- **Hiroshi G. Okuno** (for LyricSynchronizer)
- **Katunobu Itou** (for Speech Completion/Spotter)
- **Satoru Hayamizu** (for Speech Completion)
- **Koji Kitayama** (for Speech Spotter)
- **Tetsunori Kobayashi** (for Speech Spotter)
- **Tomoyasu Nakano** (for VocaListener, VocaWatcher)
- **Shuuji Kajita, Yosuke Matsusaka, Shin'ichiro Nakaoka, Yoshio Matsumoto, and Kazuhito Yokoi** (for VocaWatcher)
- **JST CrestMuse Project** (for research funding)

Please send me your comments:

E-mail: m.goto [at] aist.go.jp

URL: http://staff.aist.go.jp/m.goto/