19

Transcribing known and unknown sounds

Bayesian semi-supervised audio event recognition

Abstract— As the quantity of available multimedia data increases, a technique for automatically extracting the significant information from audio or video data becomes crucial for application to multimedia search and indexing, surveillance, and the monitoring of people in need of care. We present a novel generative model for audio event transcription that recognizes "events" on audio signals including multiple kinds of overlapping sounds. Since this model allows us to transcribe the events automatically while avoiding the model selection problem by assuming a countably infinite number of possible audio events in the input signal, we can detect unknown sounds that are not included in the training data. Experimental results show that our model is better at annotating an audio signal than a baseline method.

Recognize "events" on audio signals including multiple kinds of overlapping sounds

Ambient Singing Music Speech Applause

☹ **Difficulties**
- Some audio events are overlapped.
- Annotated databases are still sparse.
- Unknown sounds cannot be detected.

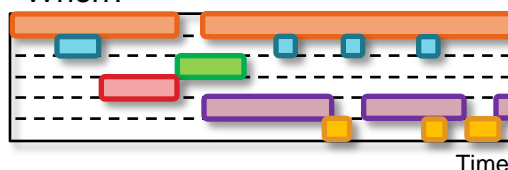## Automatically transcribe the events based on a semi-supervised learning setup

Time

Labeled data
- Music
- Sound effect
- Applause
- Speaker 1
- Speaker 2
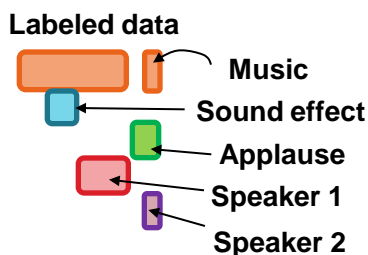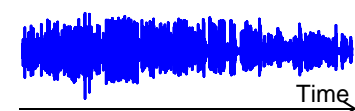
Key point:
Assuming a countably infinite number of possible audio events in the input signal
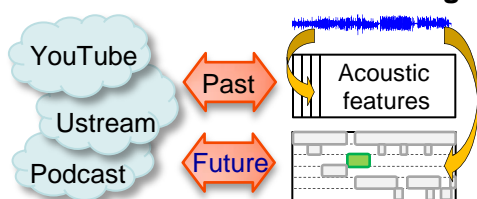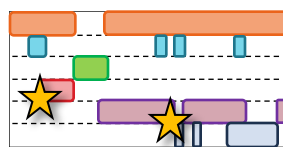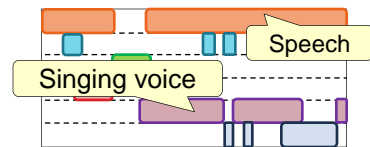
What kind of audio events?

Music   Applause   ⋯   Unknown sounds ?

When?

Time

Applications

Multimedia search and indexing

YouTube
Ustream
Podcast

Past   Future

Acoustic features

Surveillance and monitoring of people in need of care

Speech analysis and speaker diarization

Speech

Singing voice

Related works

[1] Y. Ohishi, D. Mochihashi, T. Matsui, M. Nakano, H. Kameoka, T. Izumitani, K. Kashino, "Bayesian semi-supervised audio event transcription based on Markov Indian buffet process," in *Proc. 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* 2013.
[2] Y. Ohishi, D. Mochihashi, T. Matsui, M. Nakano, H. Kameoka, T. Izumitani, K. Kashino, "Bayesian semi-supervised audio event diarization based on Markov Indian buffet process," in *Proc. The 2013 Spring Meeting of The Acoustical Society of Japan (Japanese edition),* 1-9-6, pp. 19-22, 2013.

Contact

Yasunori Ohishi   Media Recognition Research Group, Media Information Laboratory
E-mail : ohishi.yasunori{at}lab.ntt.co.jp (Please replace {at} with @)