

## Conference schedule

### 30th September

Registration and welcoming dinner

### 1st October

9:00 – 9:10      Opening remarks

9:10 – 10:10      Keynote

Robot as a multimodal human interface device  
*Tetsunori Kobayashi*

10:10 – 10:20      Coffee break

10:20 – 11:40      Session 1: Recognition

Cascading appearance-based features for visual voice activity detection  
*Rajitha Navarathna, David Dean, Patrick Lucey, Sridha Sridharan and Clinton Fookes*

Audio-visual speech recognition system for a robot  
*Takami Yoshida and Kazuhiro Nakadai*

Audio-visual television broadcast programs processing, transcription, indexing and searching  
*Josef Chaloupka and Jan Nouza*

Decision fusion by boosting method for multi-modal voice activity detection  
*Shin'ichi Takeuchi, Takashi Hashiba, Satoshi Tamura and Satoru Hayamizu*

11:40 – 13:20      Lunch

13:20 – 14:20      Session 2: Perception - McGurk effect -

Disentangling unisensory from fusion effects in the attentional modulation of McGurk effects:  
a Bayesian modeling study suggests that fusion is attention-dependent  
*Jean-Luc Schwartz, Kaisa Tiippana and Tobias Andersen*

Is there a McGurk effect for tongue reading?  
*Olov Engwall*

The McGurk illusion in the oddity task  
*Tobias S. Andersen*

14:20 – 14:30      Coffee break

14:30 – 15:30      Session 3: Emotion, Prosody

Abstracting visual prosody across speakers and face areas  
*Erin Cvejic, Jeusun Kim and Chris Davis*

Emotion perception by eye and ear and halves and wholes  
*Jeusun Kim and Chris Davis*

Cross-cultural differences in the multisensory perception of emotion  
*Akihiro Tanaka, Ai Koizumi, Hisato Imai, Saori Hiramatsu, Eriko Hiramoto and Beatrice de Gelder*

15:30 – 15:40      Comfort break

15:40 – 16:20 Session 4: Perception

Long-term cochlear implant users have resistance to noise, but short-term users don't

*Yori Kanekama, Satoko Hisanaga, Kaoru Sekiyama, Narihiro Kodama, Yasuhiro Samejima, Takao Yamada and Eiji Yumoto*

Production of Mandarin lexical tones: Auditory and visual components

*Virginie Attina, Guillaume Gibert, Eric Vatikiotis-Bateson and Denis Burnham*

16:20 – 16:30 Comfort break

16:30 – 18:00 Posters: Recognition, Synthesis

Limitations of visual speech recognition

*Jacob L. Newman, Barry-John Theobald and Stephen J. Cox*

Investigating the role of the Lombard reflex in visual and audiovisual speech recognition

*Panikos Heracleous, Miki Sato, Carlos T. Ishi, and Norihiro Hagita*

Acoustic head gesture recognition and its applications

*Akira Sasou, Yasuharu Hashimoto and Katsuhiko Sakaue*

Evaluation of real-time audio-visual speech recognition

*Peng Shen, Satoshi Tamura and Satoru Hayamizu*

Real-time audio-visual voice activity detection for speech recognition in noisy environments

*Carlos T. Ishi, Miki Sato, Norihiro Hagita and Shihong Lao*

CENSREC-1-AV: An audio-visual corpus for noisy bimodal speech recognition

*Satoshi Tamura, Chiyomi Miyajima, Norihide Kitaoka, Takeshi Yamada, Satoru Tsuge, Tetsuya Takiguchi, Kazumasa Yamamoto, Takanobu Nishiura, Masato Nakayama, Yuki Denda, Masakiyo Fujimoto, Shigeki Matsuda, Tetsuji Ogawa, Shingo Kuroiwa, Kazuya Takeda and Satoshi Nakamura*

Exploring visual features through gabor representations for facial expression detection

*Sien W. Chew, Patrick Lucey, Sridha Sridharan and Clinton Fookes*

Towards a true acoustic-visual speech synthesis

*Asterios Toutios, Utpala Musti, Slim Ouni, Vincent Colotte, Brigitte Wrobel-Dautcourt and Marie-Odile Berger*

Building speaker-specific lip models for talking heads from 3D face data

*Takaaki Kuratate and Marcia Riley*

19:00 – Conference dinner

**2nd October**

9:00 – 10:00 Keynote

What do human infants expect when adults communicate to them?

*Gergely Csibra*

10:00 – 10:10 Coffee break

10:10 – 11:10 Session 5: Perception - Brain -

Brain regions differentially involved with multisensory and visual only speech gesture information

*Daniel E. Callan*

Impact of language on audiovisual speech perception examined by fMRI  
*Jun Shinozaki, Kaoru Sekiyama, Nobuo Hiroe, Taku Yoshioka and Masaaki Sato*

An ERP examination of audiovisual speech perception in Japanese younger and older adults  
*Satoko Hisanaga, Kaoru Sekiyama, Tomohiko Igasaki and Nobuki Murayama*

11:10 – 11:20    Comfort break

11:20 – 12:00    Session 6: Perception - Infants -

Infants match auditory and visual speech in schematic point-light displays  
*Christine Kitamura and Jeesun Kim*

I can see what you said: Infant sensitivity to articulator congruency between audio-only and silent-video presentations of native and nonnative consonants  
*Catherine T. Best, Christian Kroos and Julia Irwin*

12:00 – 14:00    Lunch

14:00 – 15:00    Session 7: Recognition

A study of influence of word lip reading by change of frame rate  
*Takeshi Saitoh and Ryosuke Konishi*

Detection of specific mispronunciations using audiovisual features  
*Sébastien Picard, G. Ananthakrishnan, Preben Wik, Olov Engwall and Sherif Abdou*

Improving visual features for lip-reading  
*Yuxuan Lan, Barry-John Theobald, Richard Harvey, Eng-Jon Ong and Richard Bowden*

15:00 – 15:10    Comfort break

15:10 – 16:10    Session 8: Synthesis

Optimized photorealistic audiovisual speech synthesis using active appearance modeling  
*Wesley Mattheyses, Lukas Latacz and Werner Verhelst*

In pursuit of visemes  
*Sarah Hilder, Barry-John Theobald and Richard Harvey*

Acoustic-to-articulatory inversion in speech based on statistical models  
*Atef Ben Youssef, Pierre Badin and Gérard Bailly*

16:10 – 16:20    Comfort break

16:20 – 17:50    Posters: Perception, Emotion, Interaction

The McGurk effect at various auditory signal-to-noise ratios in American and Finnish listeners  
*Kaisa Tiippana, Erin Hayes, Riikka Möttönen, Nina Kraus and Mikko Sams*

Binding and unbinding in audiovisual speech fusion: *Removing the McGurk effect by an incoherent preceding audiovisual context*  
*Olha Nahorna, Frédéric Berthommier and Jean-Luc Schwartz*

Role of form and motion information in auditory-visual speech perception of McGurk combinations and fusions  
*Guillaume Gibert, Andrew Fordyce and Catherine J. Stevens*

Speech-specificity of two audiovisual integration effects  
*Kasper Eskelund, Jyrki Tuomainen and Tobias S. Andersen*

The multimodal analysis for understanding child behavior focused on attention-catching  
*Shogo Ishikawa, Shinya Kiriyaama, Yoichi Takebayashi and Shigeyoshi Kitazawa*

A study of speech interface for living space adapting to user environment by considering  
scenery situation

*Kenichi Shibata, Shinya Kiriyaama, Tomohiro Haraikawa, Yoichi Takebayashi and  
Shigeyoshi Kitazawa*

Effects of speech-rate conversion on asynchrony perception of audio-visual speech

*Shiho Miyazawa, Akihiro Tanaka, Shuichi Sakamoto and Takehiko Nishimoto*

The effects of anxiety on the perception of emotion in the face and voice

*Ai Koizumi, Akihiro Tanaka, Hisato Imai, Saori Hiramatsu, Eriko Hiramoto, Takao Sato  
and Beatrice de Gelder*

“D-o-e-s-Not-C-o-m-p-u-t-e”: Vowel hyperarticulation in speech to an auditory-visual avatar

*Denis Burnham, Sebastian Joeffry and Lauren Rice*

17:50 – 18:00    Closing remarks

**3rd October**

Breakfast and departure