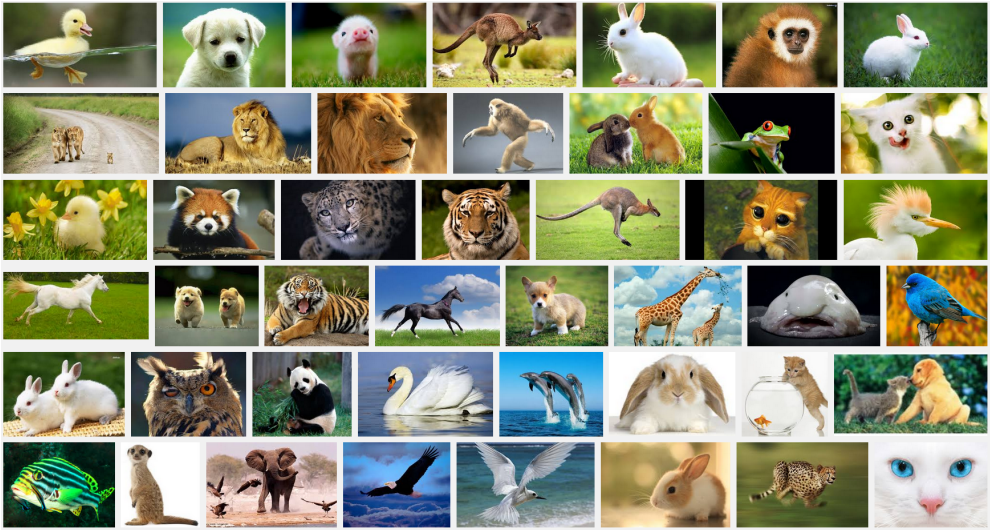
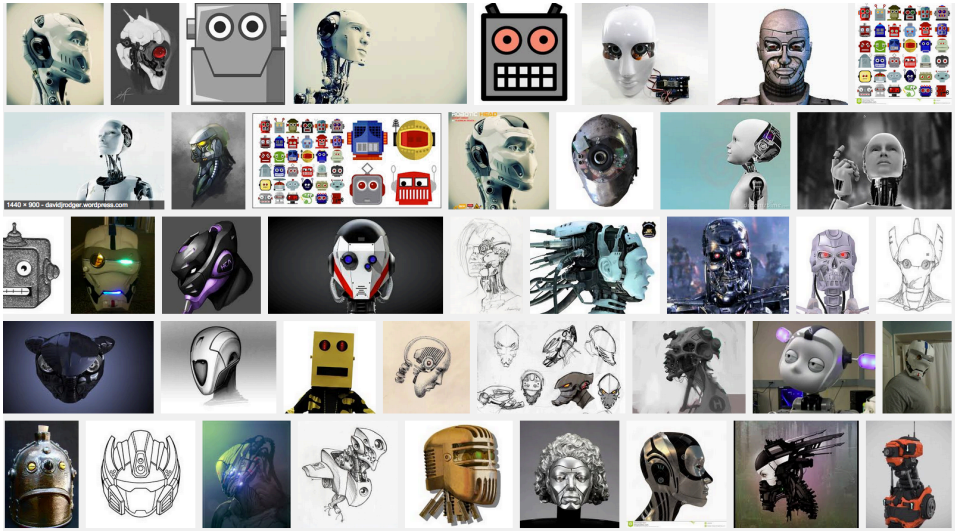
5. Fusion of Audio and Vision

1. Audio-visual processing challenges
2. Representation of visual information
3. The geometry of vision
4. Audio-visual feature association
5. Audio-visual alignment
6. Visually-guided audio localization
7. Audio-visual event localization
8. Audio-visual clustering
9. **Conclusions**

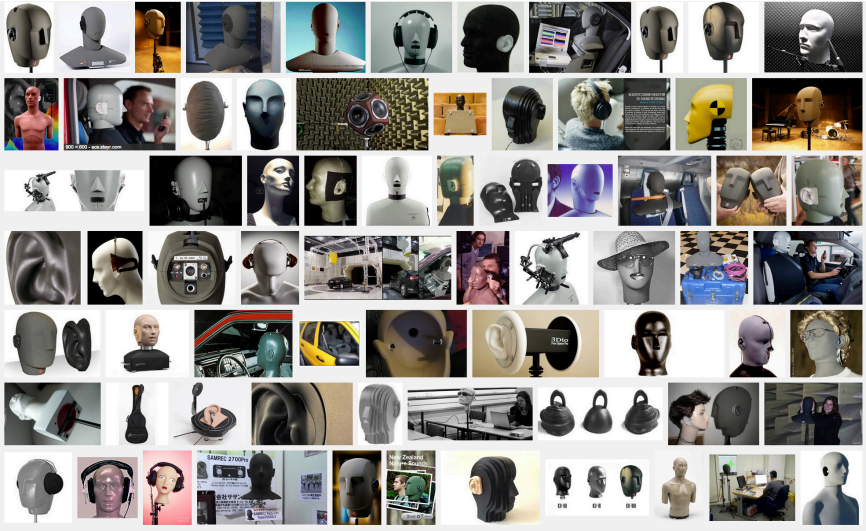
Binaural Hearing in Nature



Robot Heads



Binaural Acoustic Heads



HCI versus HRI

- Hearing allows interaction with highly rich content.
- Human-computer interaction is restricted to a single user.
- Human-robot interaction enables unconstrained multimodal communication between people and robots.

Audio Signal Processing

- Discrete auditory signals
- Discrete short-time Fourier transform
- Audio Spectrograms
- Binaural feature extraction

Sound-Source Localization

- Direct propagation model
- Geometry of two or more microphones
- Implicit propagation models
- Learning a propagation model

Machine Learning for Binaural Hearing

- Binaural features for localization
- Supervised and unsupervised learning methods
- Manifold learning
- Piecewise linear regression
- Supervised sound-source separation and localization

Audio-Visual Analysis

- Audio-visual feature association
- Camera models and Camera-microphone setups
- Audio-visual alignment
- Audio-visual event localization

Acknowledgements

- Xavier Alameda-Pineda,
- Soraya Arias,
- Antoine Deleforge,
- Vincent Drouard,
- Florence Forbes,
- Sharon Gannot,
- Israel-Dejene Gebru,
- Laurent Girin,
- Xiaofei Li,
- Jordi Sanchez-Rieira,
- and many others...

Sponsorship

- EU FP7 project HUMAVIPS
  `http://humavips.inrialpes.fr/`
- EU FP7 project EARS
  `http://robot-ears.eu/`
- ERC Advanced Grant VHIA
  `https://team.inria.fr/perception/vhia/`
- Aldebaran Robotics