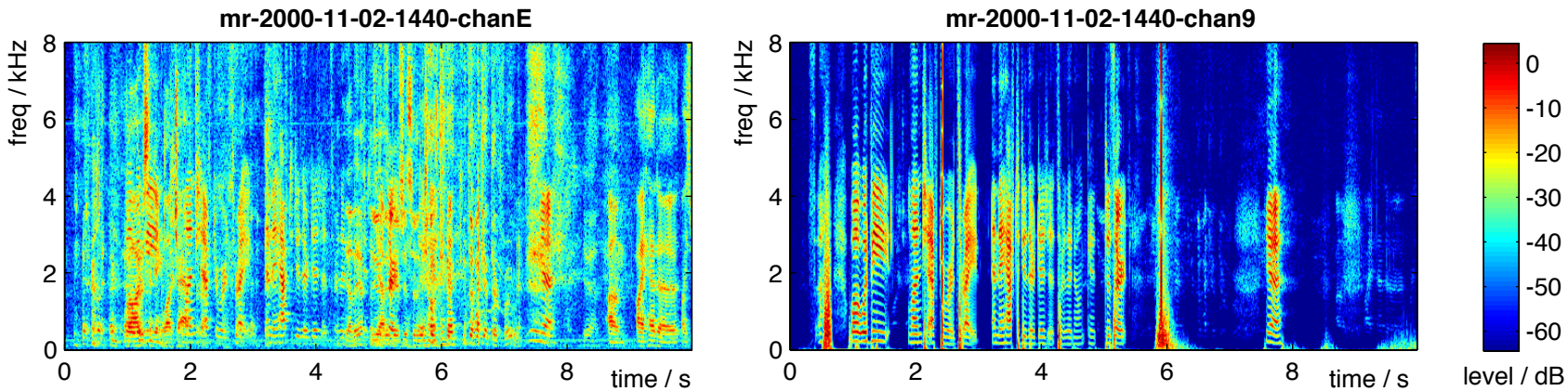
Enhancing the Intelligibility of Speech in Speech Noise

Dan Ellis (Columbia)
Pierre Divenyi (EBIRE)
Alain de Cheveigné (ENS Paris)
Te-Won Lee (UCSD)
Barbara Shinn-Cunningham (BU)
DeLiang Wang (OSU)

1. Speech in Speech Noise Problem
2. Approaches and Goals
3. The Team

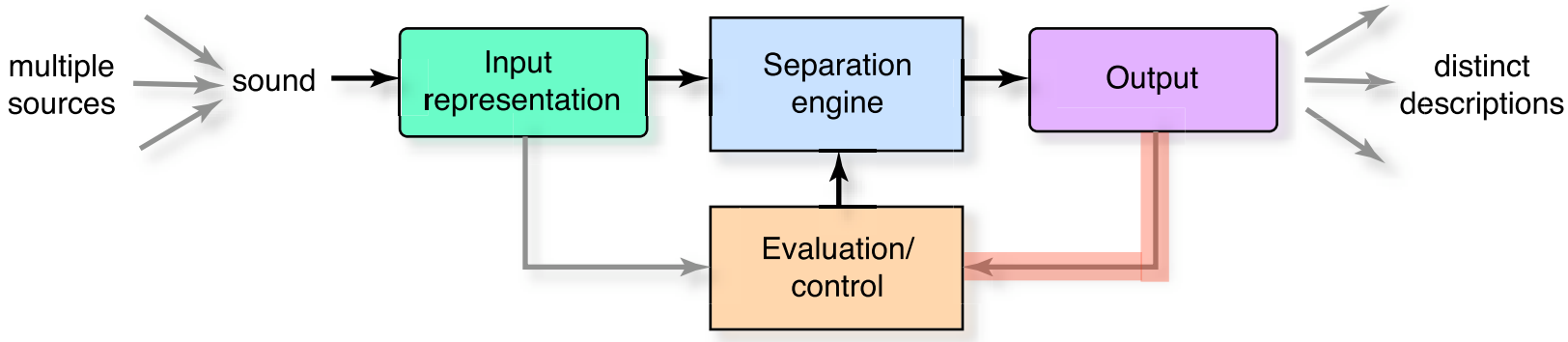
1. Speech in Speech Noise

- Cocktail-party problem: Unintelligible voices



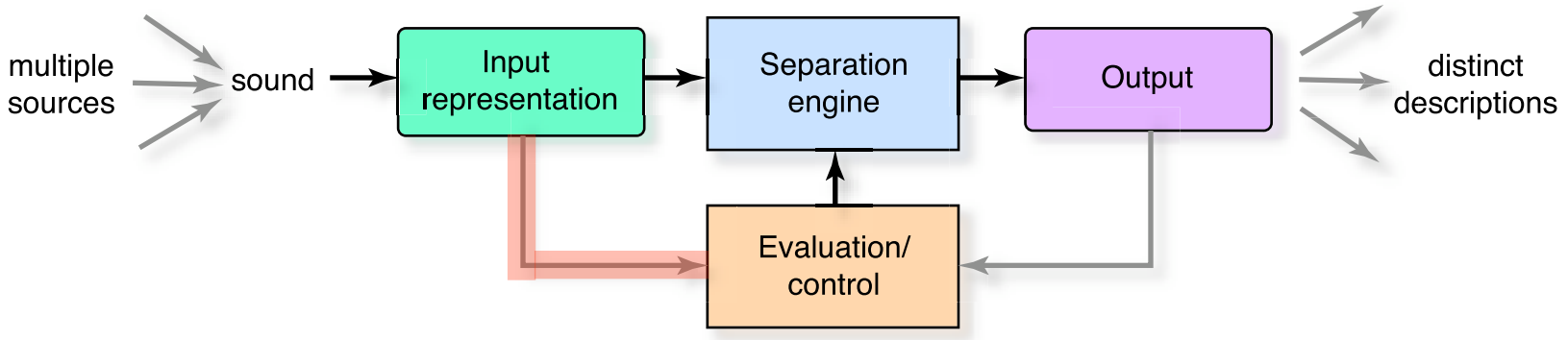
- Scenarios:
  - real-world speech noise - crowds, reverb, etc.
  - improve intelligibility
- Applications
  - hearing prostheses, communication devices
  - audio archive review, surveillance

2. Acoustic Source Separation



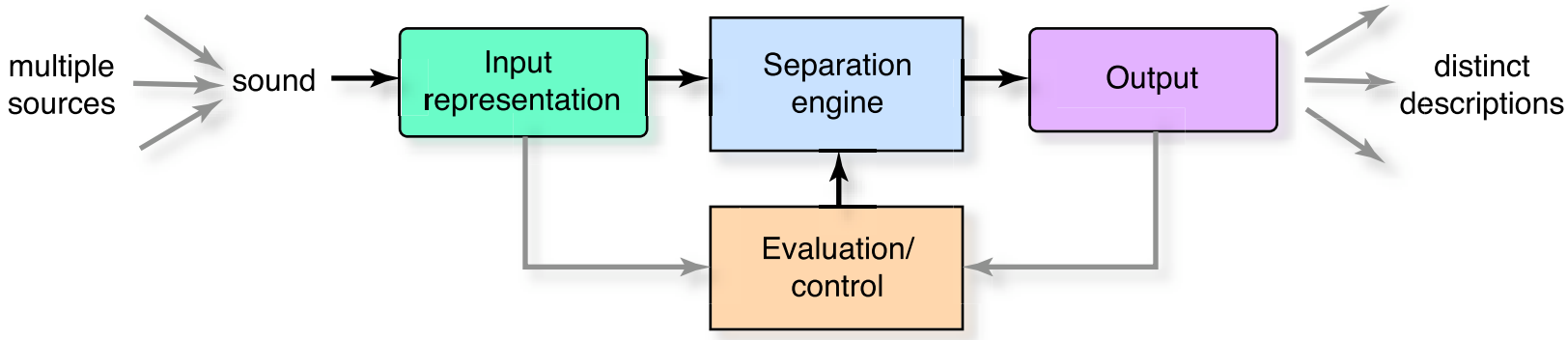
- ICA (Bell & Sejnowski '95 et seq.):
  - Input: waveform (or STFT)
  - Output: waveform (or STFT)
  - Engine: cancellation
  - Control: statistical independence of outputs
    - or energy minimization for beamforming

2. Acoustic Source Separation



- ICA (Bell & Sejnowski '95 et seq.):
- **CASA (e.g. Brown '92):**
  - Input: Periodicity, continuity, onset "maps"
  - Output: Waveform (or mask)
  - Engine: Time-frequency masking
  - Control: "Grouping cues" from input
    - or: spatial features (Roman, ...)

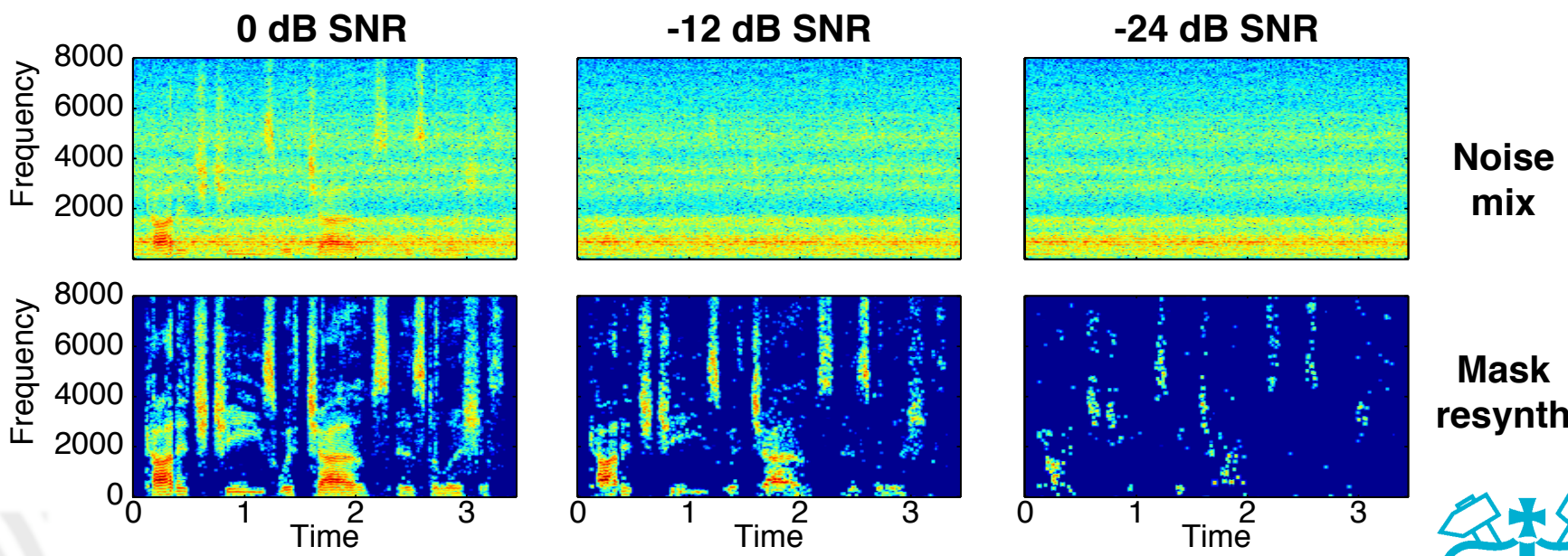
2. Acoustic Source Separation



- ICA (Bell & Sejnowski '95 et seq.):
- CASA (e.g. Brown '92):
- **Human Listeners:**
  - Input: excitation patterns ...
  - Output: percepts ...
  - Engine: ?
  - Control: find a plausible explanation

Separation Outputs

- What is the output of a separation system?
  - waveform with identified target energy
  - abstract description of content
  - reconstruction optimized for intelligibility...
- e.g. time-frequency masking



Project Goals

- Developing source separation techniques
  - single/multi channel
  - auditory/blind/model-based
  - combinations
- Collection and simulation of data
  - real-world scenarios and replicas
  - .. for parametric testing
  - .. for systematic evaluation
- Connecting with perception
  - intelligibility impact of different artifacts
  - add "proxy noise" to leverage restoration

3. A Multidisciplinary Project

Emphasis: M = machine, H = human

- Dan Ellis (Director), Organizer of Curriculum - *M*
  - machine learning, machine separation, natural scenes
- Pierre Divenyi (Co-Director), Coordination of Project Components - *H*
  - auditory scene analysis, psychoacoustics, testing methods
- Alain de Cheveigné - *H,M*
  - auditory models of separation, pitch, multichannel analysis
- Te-Won Lee - *M*
  - blind signal separation methods
- Barbara Shinn-Cunningham - *H*
  - spatial auditory scene analysis, spatial acoustics
- DeLiang Wang - *M*
  - computational auditory scene analysis, sequential/spatial grouping