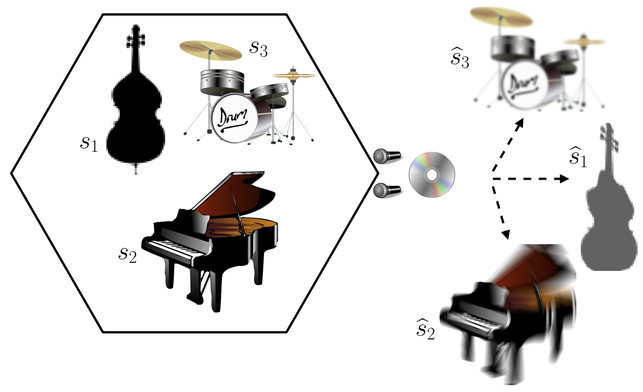
The PEASS Toolkit - Perceptual Evaluation methods for Audio Source Separation

Valentin Emiya, Emmanuel Vincent, Niklas Harlander, Volker Hohmann

The PEASS Toolkit - Perceptual Evaluation methods for Audio Source Separation

Valentin EMIYA, Emmanuel VINCENT • METISS, INRIA Rennes - Bretagne Atlantique (France)
Niklas HARLANDER, Volker HOHMANN • Carl von Ossietzky Universität Oldenburg (Germany)

THE PEASS TOOLKIT: overview



A toolkit for the perceptual evaluation of audio source separation

- The PEASS Software: a set of objective measures to predict the perceptual quality of the source/image estimates
- The PEASS Listening Test GUI: a Matlab MUSHRA GUI realized for the proposed test protocol
- The PEASS Subjective Database: a set of subjective measures resulting from listening tests (20 subjects $\times$ 80 sounds $\times$ 4 rating criteria)

The PEASS Toolkit is freely available at
http://bass-db.gforge.inria.fr/peass/

[1] **Subjective and objective quality assessment of audio source separation,** *V. Emiya, E. Vincent, N. Harlander, V. Hohmann,* IEEE Trans. on Audio, Speech and Language Processing, submitted, 2010.

[2] **Multi-criteria subjective and objective evaluation of audio source separation,** *V. Emiya, E. Vincent, N. Harlander, V. Hohmann,* AES 38th Int. Conf. on Sound Quality Evaluation, Pitea, Sweden, June 2010.

MOTIVATION: the need for a multi-criteria perceptually-based evaluation

Existing model for distortion decomposition [3]:

$$\widehat{s}_j(t) - s_j(t) = e_j^{\text{target}}(t) + e_j^{\text{interf}}(t) + e_j^{\text{artif}}(t) \quad (1)$$

- e_j^{target} denotes the error component related to the target distortion,
- e_j^{interf} denotes the interference from concurrent sources,
- e_j^{artif} is the remaining distortion component (artifacts and noise).

Defining and estimating the distortion components e_j^{target}, e_j^{interf}, e_j^{artif} is not trivial. Due to the allowed distortions in use today (time-invariant spatial and filtering distortions), the decomposition is not satisfying.

Existing quality measures: energy ratios SDR, ISR, SIR, SAR are poorly correlated with subjective scores.

Proposed multi-criteria listening test protocol

A series of 4 MUSHRA tests including several dedicated anchors:

- (T_1) Rate the *global quality* compared to the reference.
- (T_2) Rate the quality in terms of *preservation of the target source*.
- (T_3) Rate the quality in terms of *suppression of other sources*.
- (T_4) Rate the quality in terms of *absence of additional artificial* noise.

PROPOSED OBJECTIVE MEASURES: what's new?

A better distortion decomposition is achieved by:

- splitting the signals into subbands using gammatone filters;
- segmenting each subband signal into overlapping frames;
- decomposing each frame into distortion components using a matched FIR filter
- reconstructing the full distortion components

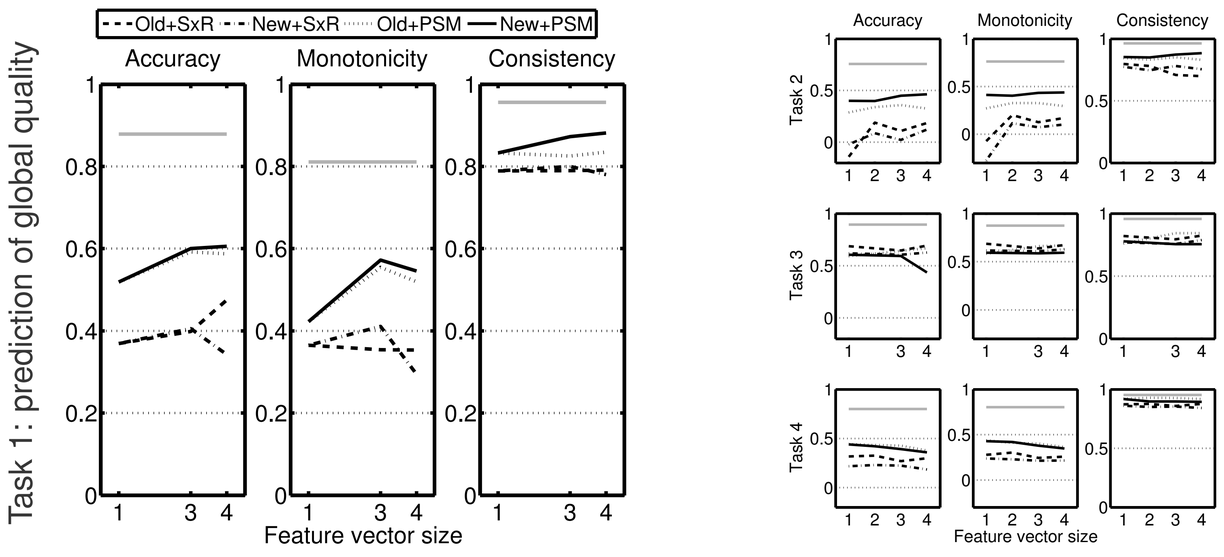
Some auditory-motivated features are derived using PEMO-Q/PSM [4]:

$$q_j^{\text{overall}} \triangleq \text{PSM}(\widehat{\mathbf{s}}_j, \mathbf{s}_j) \quad (2)$$
$$q_j^{\text{target}} \triangleq \text{PSM}(\widehat{\mathbf{s}}_j, \widehat{\mathbf{s}}_j - \mathbf{e}_j^{\text{target}}) \quad (3)$$
$$q_j^{\text{interf}} \triangleq \text{PSM}(\widehat{\mathbf{s}}_j, \widehat{\mathbf{s}}_j - \mathbf{e}_j^{\text{interf}}) \quad (4)$$
$$q_j^{\text{artif}} \triangleq \text{PSM}(\widehat{\mathbf{s}}_j, \widehat{\mathbf{s}}_j - \mathbf{e}_j^{\text{artif}}) \quad (5)$$

By combining the 4 features in a non-linear way to predict subjective scores $(T_1) - (T_4)$, a set of objective measures is finally output:

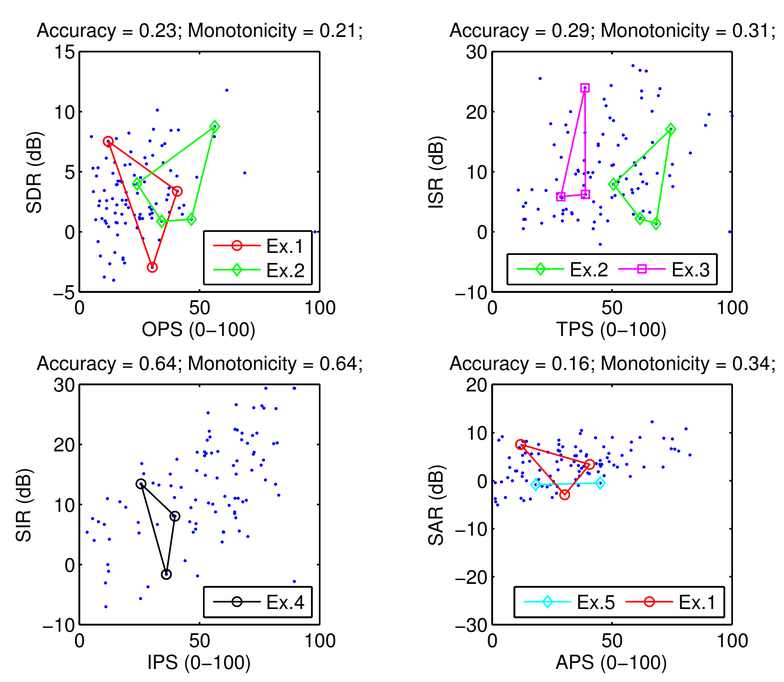
- **OPS:** the Overall Perceptual Score,
- **TPS:** the Target-related Perceptual Score,
- **IPS:** the Interference-related Perceptual Score,
- **APS:** the Artifacts-related Perceptual Score.

EVALUATION RESULTS: prediction performance and evaluation at SiSEC 2010



Prediction results (cross-validation on the PEASS database) for the 4 tasks: curves are various combinations of the old/new decompositions with the energy ratio/PEMO-Q measures, as a function of the number of features.

DEMO SiSEC 2010



BSS eval vs. PEASS: scatter plots of the SiSEC 2010 results for the set of *Professionally produced music recordings*.

[3] E. Vincent, R. Gribonval, C. Févotte, *Performance measurement in blind audio source separation*, IEEE Trans. on Acoustics, Speech and Signal Proc., 14 (4), 2006.

[4] R. Huber, B. Kollmeier, *PEMO-Q – A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception*, IEEE Trans. on Acoustics, Speech and Signal Proc., 14 (6), 2006.

9th Int. Conf. on Latent Variable Analysis and Signal Separation, September 27-30, 2010, St. Malo, France