

Audio inpainting: problem statement, relation with sparse representations and some experiments

Amir Adler, Valentin Emiya, Maria Jafari, Michael Elad, Rémi Gribonval,
Mark Plumbley

► **To cite this version:**

Amir Adler, Valentin Emiya, Maria Jafari, Michael Elad, Rémi Gribonval, et al.. Audio inpainting: problem statement, relation with sparse representations and some experiments. SMALL Workshop on Sparse Dictionary Learning, Jan 2011, London, United Kingdom. 2011. <inria-00560110>

HAL Id: inria-00560110

<https://hal.inria.fr/inria-00560110>

Submitted on 7 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Audio inpainting: problem statement, relation with sparse representations and some experiments

Amir ADLER¹, Valentin EMIYA², Maria JAFARI³, Michael ELAD¹, Rémi GRIBONVAL², Mark PLUMBLEY³

¹ The Technion, Israel ² INRIA Rennes - Bretagne Atlantique, France ³ Queen Mary University of London, UK

PROBLEM STATEMENT & APPLICATIONS: a unified scheme for existing tasks

Audio inpainting scheme: an inverse problem where

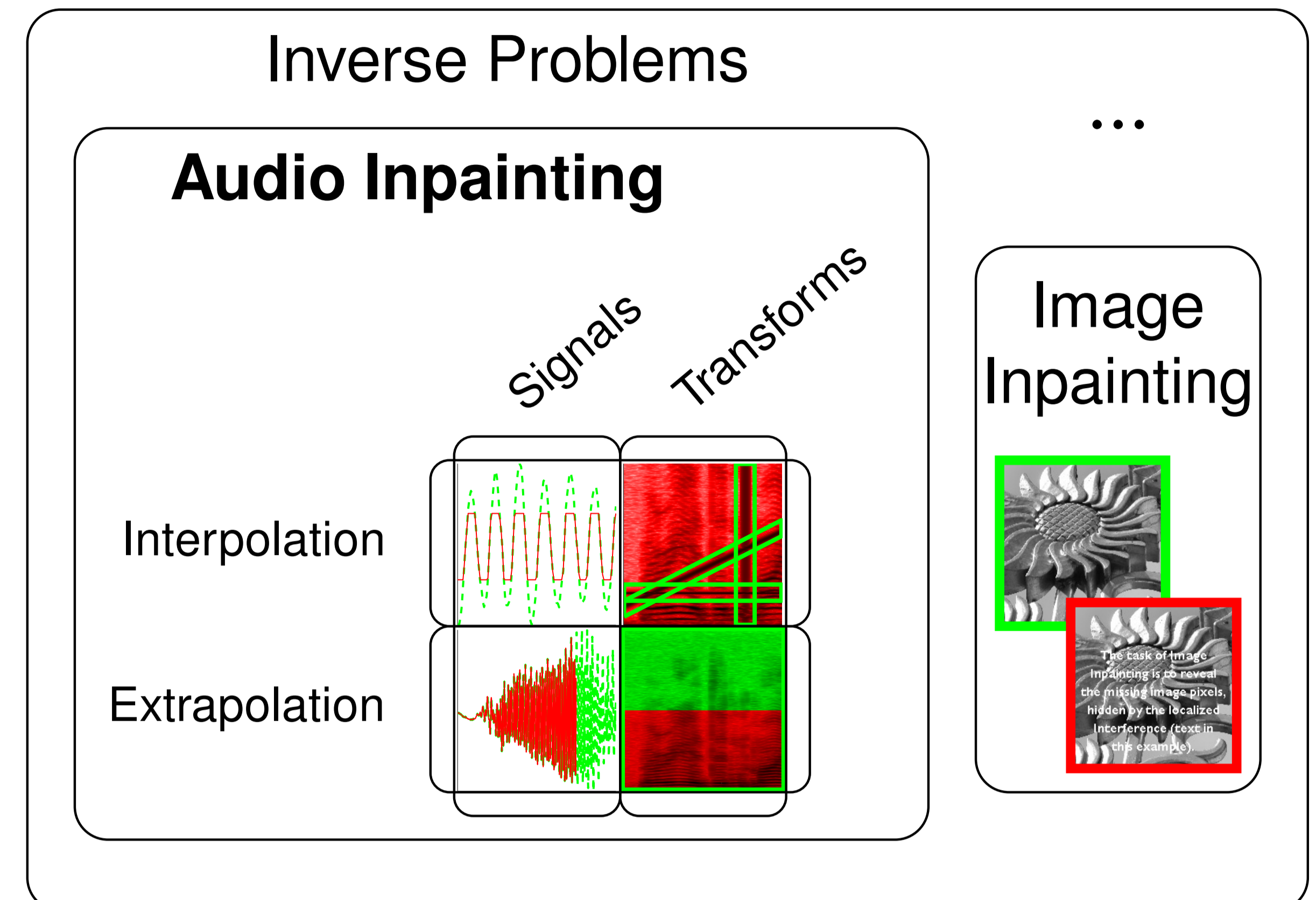
- a set of reliable audio data \mathbf{y} is observed,
 - one must estimate the remaining missing or highly corrupted data.
- A unified scheme covering existing subproblems referred to as interpolation, extrapolation, imputation, (bandwidth) extension.

Formulation: estimate the original data \mathbf{s} from \mathbf{y} given \mathbf{M}

$$(1) \mathbf{y} = \mathbf{M}\mathbf{s} \text{ with } \begin{cases} \mathbf{s} & \in \mathbb{R}^N \\ \mathbf{y} & \in \mathbb{R}^{N-M} \\ \mathbf{M} & \in \mathbb{R}^{(N-M) \times N} \end{cases} \text{ and } \begin{matrix} \text{---} & \mathbf{M} \\ \text{---} & \text{---} \\ \text{---} & \text{---} \\ \text{---} & \text{---} \\ \text{---} & \text{---} \\ \text{---} & \text{---} \\ \text{---} & \text{---} \\ \text{---} & \text{---} \\ \text{---} & \text{---} \\ \text{---} & \text{---} \end{matrix}$$

Several kinds of audio data: waveforms [1,2], transforms or mid-level representations [3].

A number of applications: removing clicks in old recordings, declipping, packet loss concealment in VoIP or P2P networks, bandwidth extension, recovery of TF coefficients masked by interfering sources/noise.



PROPOSED APPROACH: audio inpainting using sparse representations

Since the problem (1) is ill-posed, additional *a priori* is required.

Sparsity assumption on audio data

$$(2) \mathbf{s} = \mathbf{D}\mathbf{x} \quad \text{with } \begin{cases} \mathbf{D} \in \mathbb{R}^N \times \mathbb{R}^{K_D} \text{ (dictionary)} \\ \mathbf{x} \in \mathbb{R}^{K_D} \text{ (sparse coefficients)} \\ \|\mathbf{x}\|_0 \ll N \leq K_D \end{cases}$$

Noiseless ideal estimation

$$(3) \hat{\mathbf{x}} \triangleq \arg \min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ s.t. } \mathbf{y} = \mathbf{M}\mathbf{D}\mathbf{x}$$

Inpainting algorithms

Algorithm 1 Inpainting with l_1 -minimization (L_1)

Using convex minimization, do

$$\hat{\mathbf{x}} \leftarrow \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ s.t. } \|\mathbf{y} - \mathbf{M}\mathbf{D}\mathbf{x}\|_2^2 \leq \theta^2$$

$$\hat{\mathbf{s}} \leftarrow \mathbf{D}\hat{\mathbf{x}}$$

Algorithm 2 OMP-based inpainting (OMP)

Dictionary $\tilde{\mathbf{D}} = [\tilde{\mathbf{d}}_1, \dots, \tilde{\mathbf{d}}_{K_D}] \leftarrow \mathbf{M} \times \mathbf{D} \times \mathbf{W}_{\text{MD}}^{-1}$ (\mathbf{W}_{MD} : diag. matrix of norms of MD columns)

Residual $\mathbf{r}_0 \leftarrow \mathbf{y}$

Iteration counter $k \leftarrow 1$, support set $\Omega_0 \leftarrow \emptyset$

while $k \leq K_{\text{max}}$ AND $\|\mathbf{r}_k\|_2 \geq \theta_{\text{OMP}}$ **do**

Select atom: $j \leftarrow \arg \max_j |\langle \mathbf{r}_{k-1}, \tilde{\mathbf{d}}_j \rangle|$

Update support $\Omega_k \leftarrow \Omega_{k-1} \cup j$

Update current solution $\mathbf{x}_k \leftarrow \arg \min_{\mathbf{x}} \|\mathbf{y} - \tilde{\mathbf{D}}_{\Omega_k} \mathbf{x}\|_2$

Update residual $\mathbf{r}_k \leftarrow \mathbf{y} - \tilde{\mathbf{D}}_{\Omega_k} \mathbf{x}_k$

Increment iteration counter $k \leftarrow k + 1$

end while

$$\hat{\mathbf{s}} \leftarrow \mathbf{D}\mathbf{W}_{\text{MD}}\mathbf{x}_k$$

Specific inpainting algorithms for restoring clipped signals

→ Constrain the restored samples to be beyond the clipping threshold θ_{clip}

Algorithm 3 (L_1 -C)

Using convex minimization, do

$$\hat{\mathbf{x}} \leftarrow \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ s.t. } \begin{cases} \|\mathbf{y} - \mathbf{M}\mathbf{D}\mathbf{x}\|_2^2 \leq \theta^2 \\ \mathbf{M}^+ \mathbf{D}\mathbf{x} \geq \theta_{\text{clip}} \\ \mathbf{M}^- \mathbf{D}\mathbf{x} \leq -\theta_{\text{clip}} \end{cases}$$

$$\hat{\mathbf{s}} \leftarrow \mathbf{D}\hat{\mathbf{x}}$$

Algorithm 4 (OMPc)

After the while loop in Alg. 1, refine \mathbf{x}_k as

$$\mathbf{x}_k \leftarrow \arg \min_{\mathbf{x}} \|\mathbf{y} - \tilde{\mathbf{D}}_{\Omega_k} \mathbf{x}\|_2 \text{ s.t. } \begin{cases} \mathbf{M}^+ \mathbf{D}\mathbf{x} \geq \theta_{\text{clip}} \\ \mathbf{M}^- \mathbf{D}\mathbf{x} \leq -\theta_{\text{clip}} \end{cases}$$

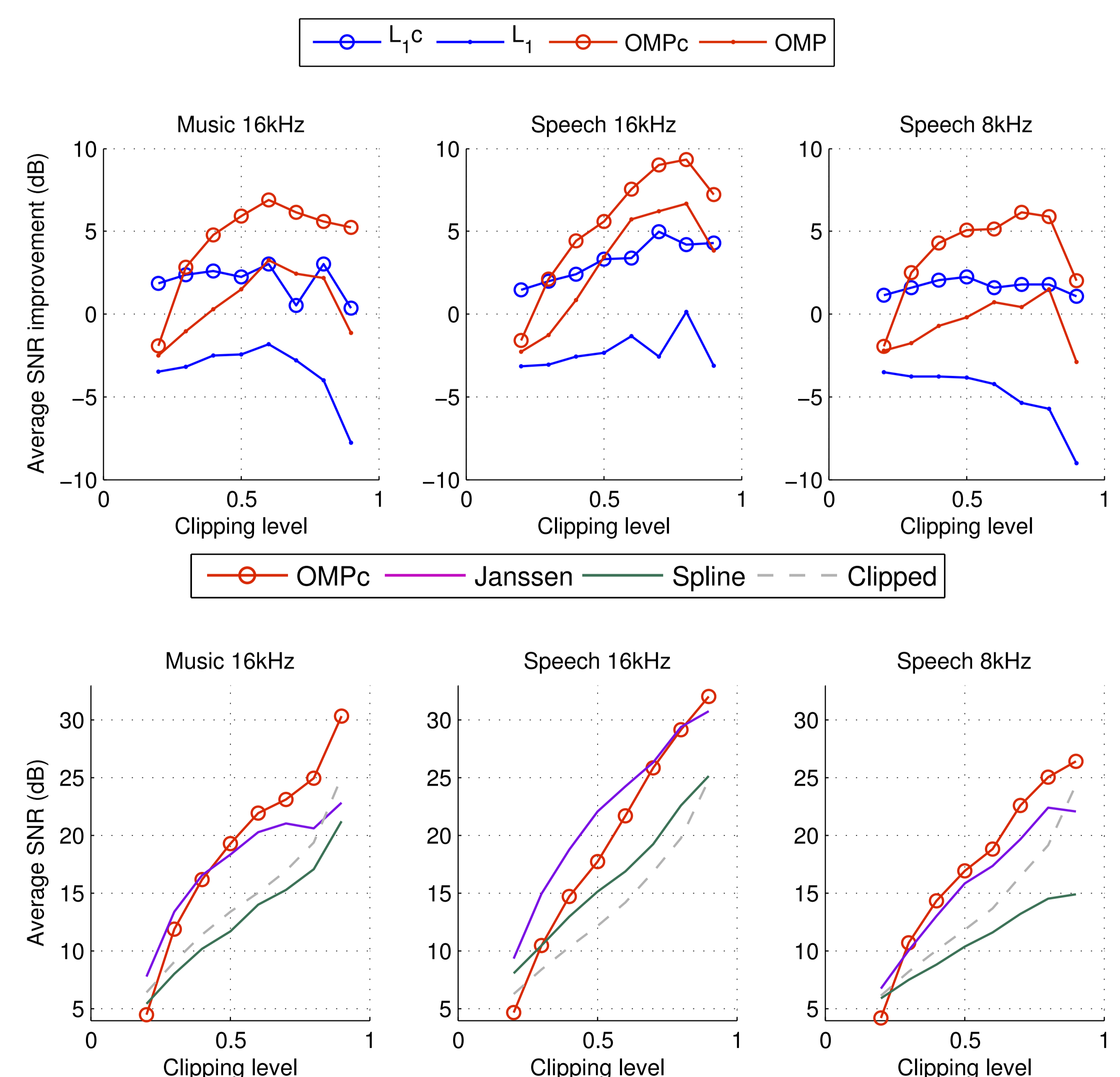
$$\hat{\mathbf{s}} \leftarrow \mathbf{D}\mathbf{W}_{\text{MD}}\mathbf{x}_k$$

where \mathbf{M}^+ and \mathbf{M}^- are the projectors onto the subspace of positive and negative clipped samples respectively.

EXPERIMENTS: declipping audio signals

Experimental settings

- 3 datasets with 10 5-seconds signals
- Frame-by-frame processing with OLA reconstruction: 75% overlap, 64ms frames, sine weighting windows
- Algorithms OMP, OMPc, L_1 , L_1 -C are applied in each frame
- OMPc is compared to spline interpolation and Janssen's algorithm [1] based on autoregressive models.
- Algorithm parameters are fixed (tuned on a separate database)
- Dictionary \mathbf{D} : redundant sine-windowed DCT
- SNR is computed on the corrupted samples only



[1] A. Janssen, R. Veldhuis, L. Vries, *Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes*, in IEEE Trans. on Acoustics, Speech and Signal Proc., 34 (2), 1986.
 [2] S. J. Godsill, P. J. W. Rayner, *Digital audio restoration - A statistical model-based approach*, Springer-Verlag London, 1998.
 [3] J. Le Roux, H. Kameoka, N. Ono, A. de Cheveigné, S. Sagayama, *Computational auditory induction by missing-data non-negative matrix factorization*, Proc. of SAPA, 2008.

