Audio inpainting: problem statement, relation with sparse representations and some experiments
Amir Adler, Valentin Emiya, Maria G. Jafari, Michael Elad, Rémi Gribonval, Mark D. Plumbley

To cite this version:

HAL Id: inria-00560110
https://inria.hal.science/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 Jafari2, 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 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 s from y given M

\[ (1) \ y = Ms \text{ with } s \in \mathbb{R}^N, y \in \mathbb{R}^{N-M}, M \in \mathbb{R}^{N-M \times N} \]

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

\[ (2) \ s = Dx \text{ with } D \in \mathbb{R}^N \times \mathbb{R}^{N-M} \text{ (dictionary)} \]
\[ x \in \mathbb{R}^N \text{ (sparse coefficients)} \]

Noiseless ideal estimation

\[ \hat{s} = \arg \min_{x} \|x\|_1 \text{ s.t. } y = MDx \]

Inpainting algorithms

\textbf{Algorithm 1:} inpainting with \( l_1 \) minimization (L1)

\[ 1. \ \hat{s} = \arg \min_{x} \|x\|_1 \text{ s.t. } \|y - MDx\|_2 \leq \epsilon \]
\[ 2. \ D \leftarrow D \hat{x} \]

\textbf{Algorithm 2:} OMP-based inpainting (OMPc)

Dictionary \( D = [d_1, \ldots, d_M] \rightarrow M \times D = W_{MD} \) ( \( W_{MD} \): diag. matrix of norms of MD columns)
Residual \( r = y - D\hat{x} \)
Iteration counter \( k = 1 \), support set \( \Omega = \emptyset \)
while \( k \leq K \) and \( \|D\cdot_{\Omega^c}r\|_2 \geq \text{ClipLevel} \) do
    Select atom \( l = \arg \max_{j} |\langle r, d_j \rangle| \text{ s.t. } |d_j|_1, |d_j|_2 > 1 \)
    Update support \( \Omega \leftarrow \Omega \cup \{l\} \)
    Update current solution \( s = \hat{s} - s |_{\Omega^c} \) \[ y - D\hat{x} \]
    Increment iteration counter \( k = k + 1 \)
end while
\[ 3. \ D \leftarrow D_{\Omega_{\text{final}}} \]

Specific inpainting algorithms for restoring clipped signals

\[ s = \arg \min_{x} \|x\|_1 \text{ s.t. } \|y - MDx\|_2 \leq \epsilon \]
\[ s \leftarrow Ds_{\Omega_{\text{final}}} \]

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, L1, L0 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 D: redundant sine-windowed DCT
- SNR is computed on the corrupted samples only