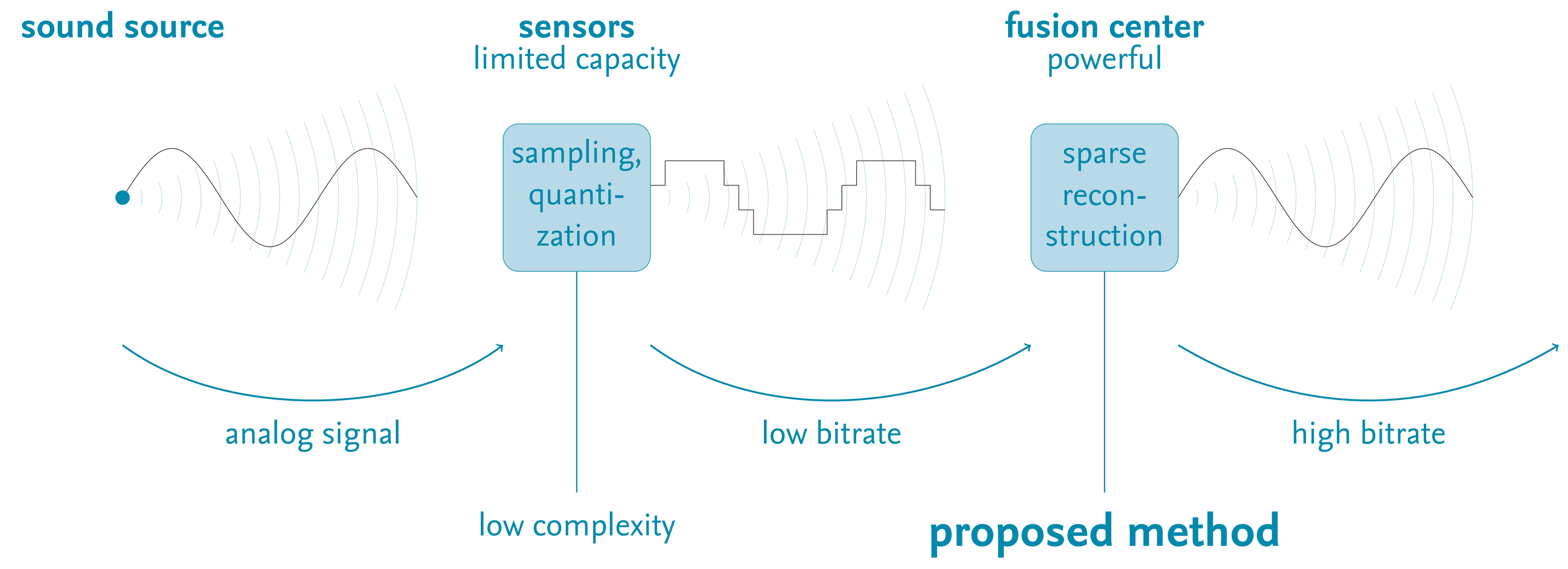
Sparse Reconstruction of Quantized Speech Signals

Christoph Brauer* Timo Gerkmann† Dirk Lorenz*

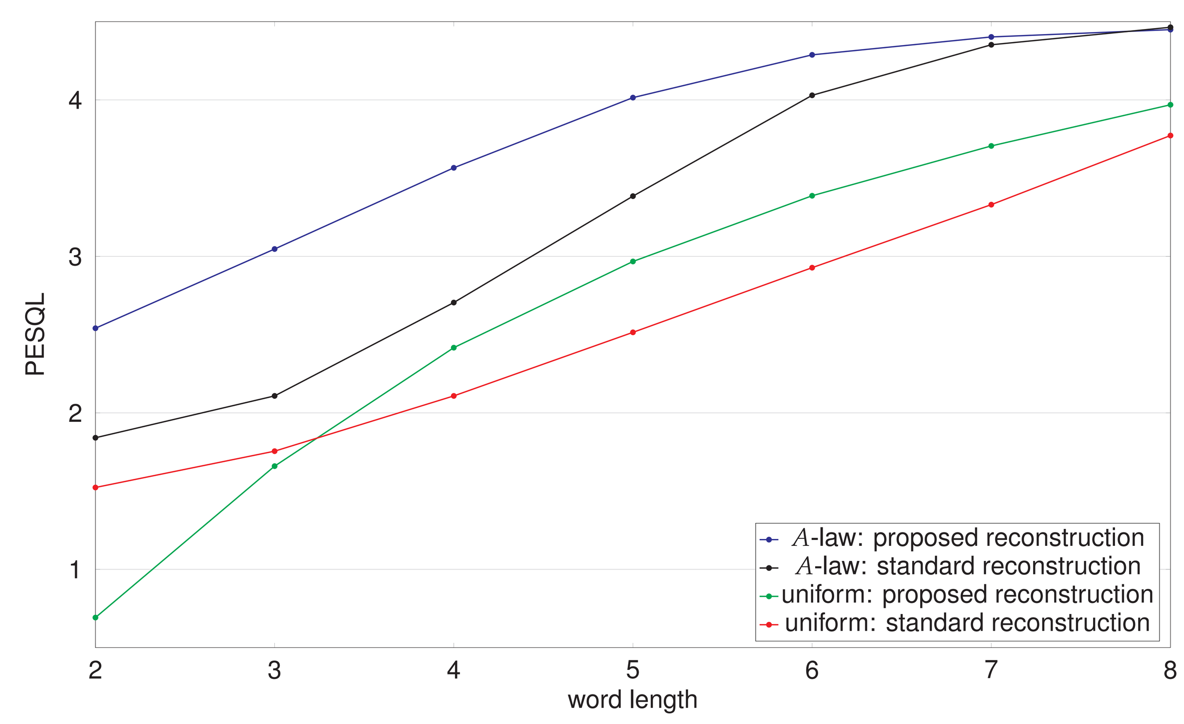
* Technische Universität Braunschweig, Germany
† Technicolor Research & Innovation, Hanover, Germany

Wireless Acoustic Sensor Networks



- At the sensor, limited power is available for coding and transmission.
- At the receiver, complex reconstruction techniques can be used.

Speech Quality Enhancement



Average PESQL values obtained in experiments with 720 speech signals from the IEEE corpus compared to average PESQL values of the associated standard reconstructions.

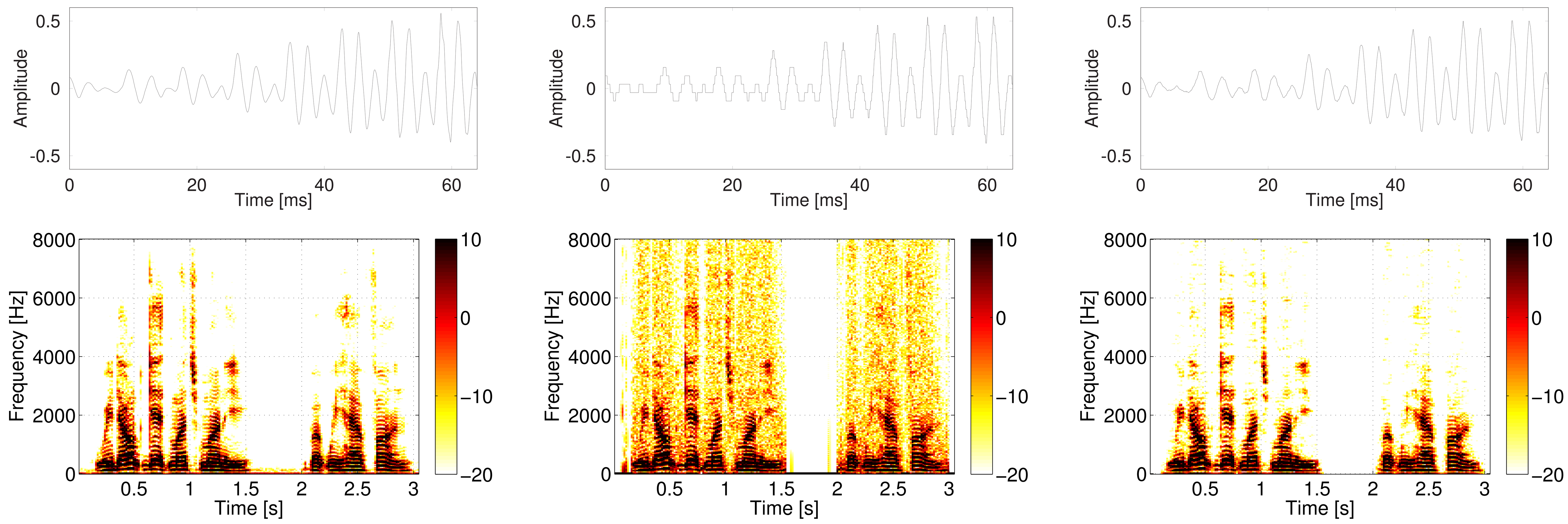
Basic Idea

- $\mathbf{f} \in \mathbf{R}^N$ is the sampled speech signal and $Q(\mathbf{f}) \in \mathbf{R}^N$ is obtained by low-bitrate scalar quantization at the sensor, e.g. ≤ 6 bit / sample.
- At this low bitrate, standard reconstruction yields poor audio quality.
- Two assumptions are crucial for our reconstruction approach:
  1. The sought after signal $\mathbf{x} \in \mathbf{R}^N$ gives the same quantized signal as the original speech signal, i.e. $Q(\mathbf{x}) = Q(\mathbf{f})$.
  2. $\mathbf{x}$ has a sparse representation in the spectral domain, i.e. $\mathbf{x} = \Psi \mathbf{a}$ for a matrix $\Psi \in \mathbf{R}^{N \times N}$ and a sparse vector $\mathbf{a} \in \mathbf{R}^N$.
- Since minimizing the ℓ_1-norm is known to support sparse solutions, our assumptions give rise to the optimization problem $\min \quad \|\mathbf{a}\|_1 \quad \text{s.t.} \quad Q(\Psi \mathbf{a}) = Q(\mathbf{f})$.

Optimization Problems and Algorithm

- Although the quantization function Q is in general non-linear, the constraint $Q(\mathbf{x}) = Q(\mathbf{f})$ has a linear reformulation:
- In case $Q = Q_\Delta$ is a uniform quantization function and the quantization intervals have length Δ, the problem turns out to be $\min \quad \|\mathbf{a}\|_1 \quad \text{s.t.} \quad \|\Psi\mathbf{a} - Q_\Delta(\mathbf{f})\|_\infty \leq \frac{\Delta}{2}$.
- In case $Q = Q_\Delta(C_A(\mathbf{f}))$ with Q_Δ as above and an A-law compression function C_A, the problem turns out to be $\min \quad \|\mathbf{a}\|_1 \quad \text{s.t.} \quad C_A^{-1}(Q_\Delta(\mathbf{f}) - \frac{\Delta}{2}) \leq \Psi\mathbf{a} \leq C_A^{-1}(Q_\Delta(\mathbf{f}) - \frac{\Delta}{2})$.
- Both are non-smooth constrained convex optimization problems.
- Numerically, we tackle both problems performing 25 iterations of the primal-dual method proposed by Chambolle and Pock.

Spectrograms



Time-domain snippets of 64 ms length (top) and spectrograms (bottom) of clean (left), quantized (middle) and reconstructed (right) speech. The snippets are taken at time 0.2s. The sampling rate is 16 kHz and the word length for the quantized speech (middle) $w = 5$ Bit.

ICASSP 2016 Shanghai, China 20-25 March 2016

technicolor

Technische Universität Braunschweig