A transient interference suppression algorithm

**Problem Formulation**

- Input signal: \( y(n) = x(n) + t(n) + d(n) \)
  - \( x(n), t(n), d(n) \) are speech, transient and (quasi) stationary noise respectively

  Speech, transient interference and stationary noise are assumed to be uncorrelated:
  \[ \lambda_s(k,l) = \lambda_x(k,l) + \lambda_t(k,l) + \lambda_d(k,l) \]

  Transients such as door knocks and keyboard tapping are short in time and widely spread across the frequency domain.

  The algorithm is based on an estimation of the transient and the enhancement of speech.

  The algorithm is robust, does not rely on transient periodicity or reoccurrence, exhibits good performance for various transient interference types.

**Algorithm Outline**

- \( y(n) = x(n) + d(n) + t(n) \)

- **Transient PSD Estimation** \( \hat{\lambda}_t(k,l) \)
  - Transient PSD estimation by computing adapted OM-LSA filter.

  Shorten STFT frames and adjust temporal recursive averaging parameter:
  \[ S(k,l) = \alpha_s S(k,l-1) + (1-\alpha_s) |Y(k,l)| \]

  Decide on transients presence according to:
  \[ S_{min}^L(k,l) = \min\{S(k,l), S(k,l-1)...S(k,l-L+1)\} \]

**Anti Causal Window For Onsets Tracking**

- Speech phonemes onsets may be wrongly considered as transients.

- Use anti-causal window (“future frames”) to distinguish between transients and speech onsets.

- Transients are assumed to be shorter than speech phonemes.

  \[ S_{min}^L(k,l) = \max\{S_{t}^L(k,l), S_{min}^{r=\infty}(k,l)\} \]

**Utilization of Signals Different Variation Rate**

- Transient are assumed to be rapidly varying compared to the slower speech and pseudo-stationary noise.

- Noise PSD estimation component of the OM-LSA is adjusted to track faster PSD change.

- Hence speech and stationary noise appear as “pseudo-stationary”.

  The adapted OM-LSA enables to estimate transient PSD.

**Experimental Results**

**Experiment setup:**

- Recorded speech signals with an average length of

- Transient instances of 40 – 140 ms durations

- Sampling rate: 16 KHz

- For PSD estimation - STFT frames of length 64, for speech enhancement - frames of length 512

- Overlap of 75% between consecutive frames

- Same maximal amplitude for speech and transients

**Speech Enhancement**

- Calculate total interference PSD estimate

  \[ \hat{\lambda}_t^*(k,l) = \hat{\lambda}_d(k,l) + \hat{\lambda}_t(k,l) \]

  Compute an optimal OM-LSA filter which minimizes:

  \[ \min\{ (\log(A(k,l)) - \log(\hat{\lambda}(k,l))) \}, A = |X(k,l)| \]