Why Frame Loss Concealment (FLC)?

- Congested Networks → Packet Loss → Audio Frame Loss

Existing FLC techniques

- Waveform repetition
- MDCT domain tonal interpolation
- Subband domain linear prediction

- X Ineffective for most audio signals as they are polyphonic.
- X Ineffective when MDCT fails to resolve individual tones.
- X Ineffective for long lost blocks.

Cascaded Long Term Prediction (CLTP)

- Predicting periodic signals such as \( x[m] = \alpha x[m - N] + \beta x[m - N_i + 1] \) is well known.
- We recently introduced CLTP for mixture of such periodic signals \( \sum x[m] \) (i.e., polyphonic signals) to improve its compression efficiency.

- Here we propose employing a CLTP synthesis filter, suitably adapted for FLC by utilizing the past and even the usually available future samples effectively.

Cascaded Long Term Prediction for Frame Loss Concealment

- Preliminary parameters are estimated from the past samples via a recursive technique.
  - Parameters of \( \hat{f} \) filter \( (1 - \alpha z^{-N - \beta z^{-N_i+1}}) \) are estimated in the residue of filtering with all the others \( (1 - \alpha z^{-N - \beta z^{-N_i+1}}) \), via the well known technique for LTP.
  - Each filter in the cascade is estimated this way in a loop until convergence.

- Using only the past samples for the filter parameter estimate doesn’t explain future samples correctly.

- So CLTP filter updated with multiplicative gain factors

\[
H_k(z) = \prod_{i=0}^{P-1} (1 - G_i(z)z^{-N_i+1} + \beta z^{N_i+1})
\]

- The gain factors are adjusted to minimize squared prediction error in the future samples.
- As cost function has complex dependency on these factors, a generic quasi-Newton optimization called L-BFGS method is used along with backtracking line search for step sizes.

- Simply predicting from past samples doesn’t ensure smooth transition into the available future samples.
- Thus lost frame samples are predicted in reverse direction from future samples with different set of CLTP gain factors.

- Final reconstruction of lost frame is a weighted average of predicted samples in each direction.

- For use in MPEG AAC, the reconstructed frame is transformed to MDCT domain and energy smoothing performed in each band \( i \), via a gain factor given as,

\[
f(i) = \left\{ \begin{array}{ll}
\sqrt{\frac{E_{\text{frame}}(i)}{E_{\text{past}}(i)}}, & \text{if } \sqrt{\frac{E_{\text{frame}}(i)}{E_{\text{past}}(i)}} > 1/T \text{ or } \frac{E_{\text{past}}(i)}{E_{\text{frame}}(i)} < 1/T, \\
1, & \text{otherwise.} 
\end{array} \right.
\]

Evaluations

- MPEG reference AAC-LD encoder used to generate 64 kbps bitstreams and the following decoders compared,
  - Reference decoder with no frame loss.
  - Reference decoder with subband domain linear prediction based FLC (SBP-FLC).
  - Reference decoder with MDCT domain tonal interpolation FLC (MDCT-FLC).
  - Reference decoder with the proposed CLTP based FLC (CLTP-FLC).

- Testing data-set: 6 audio files, 4s each, mono, 44.1/48 kHz.
- Frame loss was at the rate of 10% and random.
- Objective evaluation results of Segmental SNR in dB.

- Subjective evaluation results of MUSHRA listening tests (16 listeners, plots with average and 95% confidence interval).

Conclusions

- Currently used FLC techniques sub-optimal for polyphonic audio signals.
- Bidirectional cascaded LTP proposed for significantly improved FLC, which takes into account all the available information.
- Subjective and objective evaluations substantiate these improvements.
- Future directions include developing low complexity variant and handling burst frame losses.

http://www.scl.ece.ucsb.edu/