DELAY-LESS FREQUENCY DOMAIN PACKET-LOSS CONCEALMENT FOR TONAL AUDIO SIGNALS

Ralph Sperschneider, Janine Sukowski[1], Goran Marković

Fraunhofer

IIS

[1]now with Citrix Online GmbH, Dresden, Germany

Outline

- Motivation
- Existing solutions
- Enhanced Voice Services (EVS) Constraints
- Proposed method: Delay-less tonal Packet-Loss Concealment (PLC)
  - Integration in EVS Transform Coded Excitation (TCX) decoder
  - Phase estimation
  - Tonal component detection
- Listening test results
- Conclusions

Motivation

- Packet-switched network is used for Voice over LTE (VoLTE)
- Poor radio conditions or congestion in network → late or lost packets
- Continuity in tonal (harmonic) signals is perceptually important
  - Discontinuities are easily spotted and are unpleasant

Perfect conditions

30% of packets are lost

Existing solutions

- Time domain approaches:
  - Mute
  - Repeat frame
  - Repeat last pitch cycle
- Modified Discrete Cosine Transform (MDCT) domain approaches:
  - Domain of choice for state of the art music coding
  - Handling of lost frames:
    - Repetition with sign randomization

Perfect conditions 30% packet-loss

    - MDCT coefficient estimation for tonal components

Enhanced Voice Services (EVS) Constraints

- 32 ms delay → no additional delay is available to PLC
- 88 WMOPS → PLC complexity may not be higher than the decoding
- Limited memory consumption → low additional memory for PLC
- No degradation of intrinsic quality → no side information for PLC

Proposed method: Delay-less tonal PLC

- MDCT coefficient estimation for tonal components
  - Handles monophonic signals
  - Handles polyphonic signals
  - Handles inharmonicity
- Sign scrambling for non-tonal components
- Robust tonal component detection

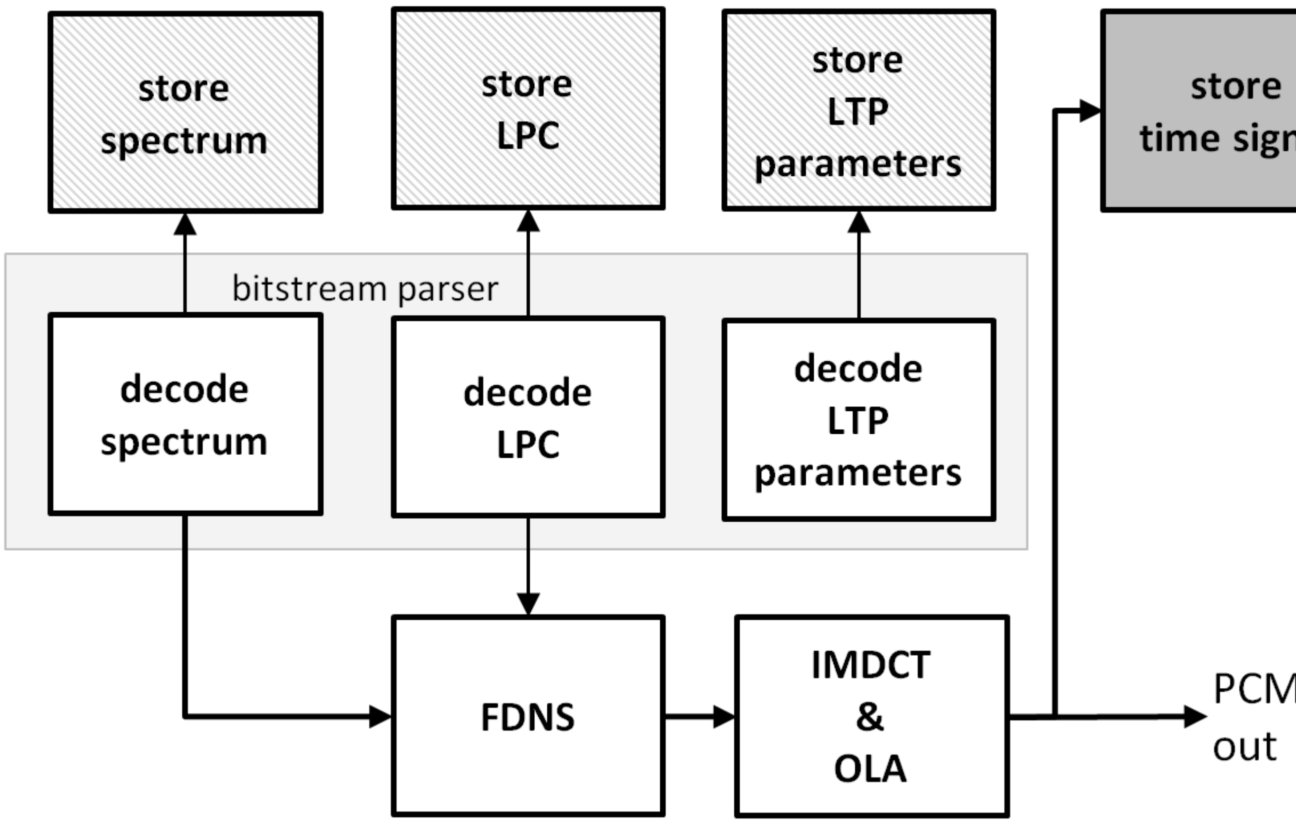
Perfect conditions

Noise substitution

No tonality detection

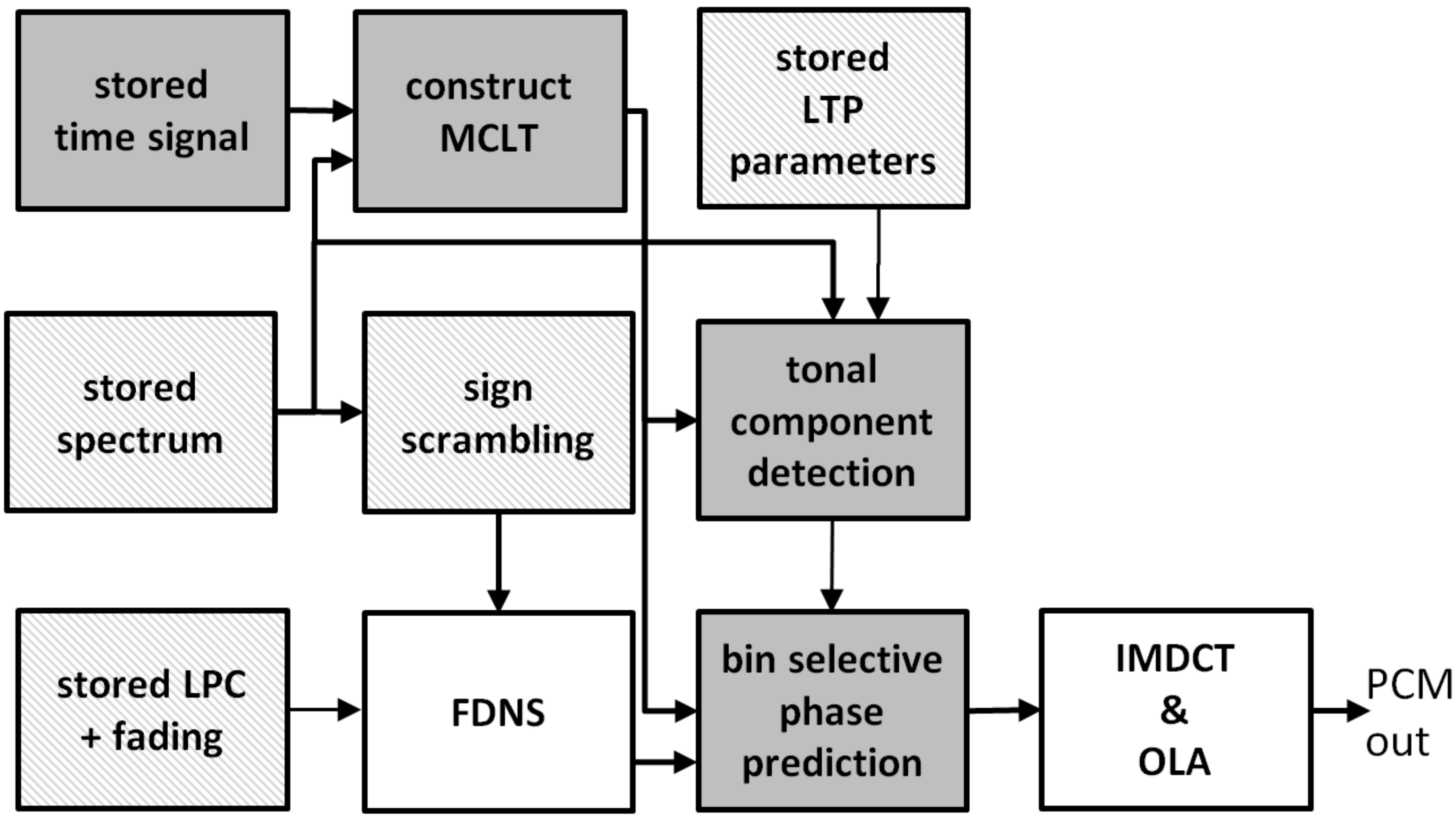
Integration in EVS TCX decoder: received frame

- EVS incorporates MDCT based Transform Coded Excitation (TCX) Codec:

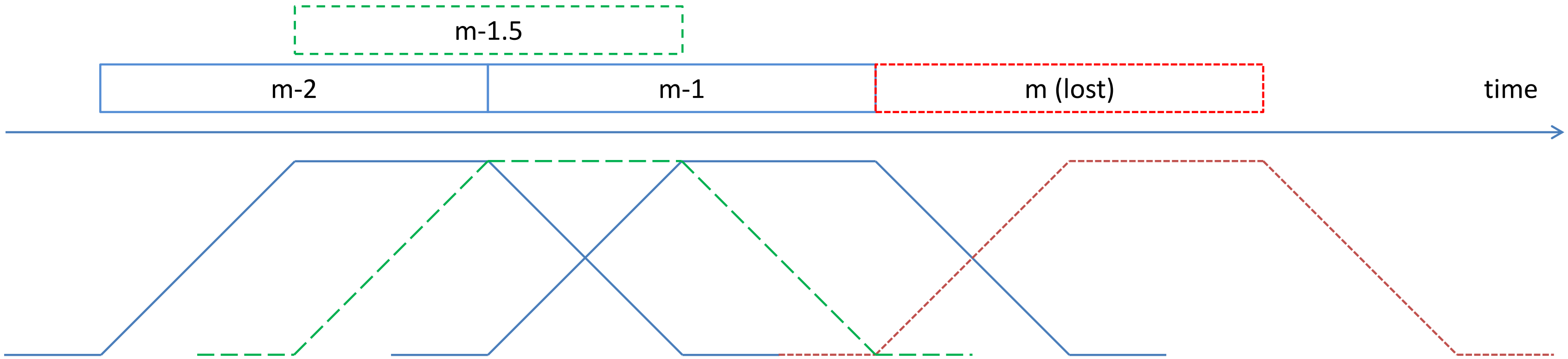


Integration in EVS TCX decoder: concealment

- Delay-less tonal PLC:



Proposed method: Delay-less tonal PLC



- MDCT coefficients: $C_{\mathrm{m-x}}(k), x \in \{1,1.5,2\}$
- MDST coefficients: $S_{\mathrm{m-x}}(k), x \in \{1,1.5,2\}$
- Modulated Complex Lapped Transform (MCLT) coefficients:

$$C_{\mathrm{m-x}}(k) + iS_{\mathrm{m-x}}(k) = Q_{\mathrm{m-x}}(k)e^{i\varphi_{\mathrm{m-x}}(k)}$$

- Magnitude: $Q_{\mathrm{m-x}}(k) = \sqrt{|S_{\mathrm{m-x}}(k)|^2 + |C_{\mathrm{m-x}}(k)|^2}$
- Phase: $\varphi_{\mathrm{m-x}}(k) = arctan\left(\frac{S_{\mathrm{m-x}}(k)}{C_{\mathrm{m-x}}(k)}\right)$

Fraunhofer IIS

Delay-less tonal PLC: Phase estimation

- $C_m(k) = Q_{\mathrm{m-x}}(k) \cdot \cos(\varphi_m(k))$
- Phase estimation: $\varphi_m(k) = \varphi_{m-x}(k) + x \cdot \Delta\varphi_{m-x,n}$
- Phase difference between successive frames:

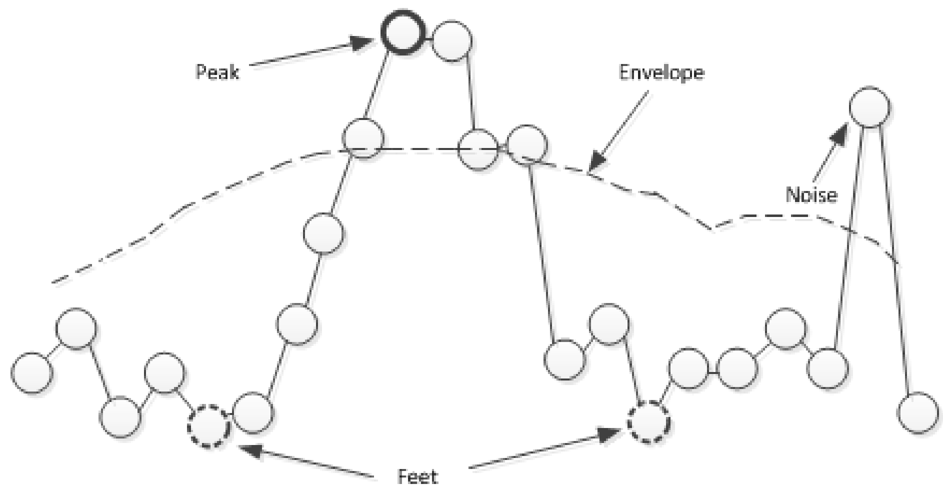
$$\Delta\varphi_{m-x,n} = \pi \cdot (l_n + \Delta l_n)$$

  - $\frac{2\pi}{M} \cdot (l_n + \Delta l_n)$ is the normalized angular frequency of the n[th] sinusoid
  - $\Delta l_n = \frac{b}{2\pi} arctan \left[\frac{cos\frac{\pi}{b} - \sqrt[G]{\frac{|Q_{\mathrm{m-x}}(l_n-1)|}{|Q_{\mathrm{m-x}}(l_n+1)|}} cos\frac{3\pi}{b}}{sin\frac{\pi}{b} + \sqrt[G]{\frac{|Q_{\mathrm{m-x}}(l_n-1)|}{|Q_{\mathrm{m-x}}(l_n+1)|}} sin\frac{3\pi}{b}} \right]$ *
  - b is the width of the windows' main lobe, G is a constant

*A. Ferreira, "Accurate estimation in the ODFT domain of the frequency, phase and magnitude of stationary sinusoids", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 47-50, 2001.

Delay-less tonal PLC: Tonal component detection

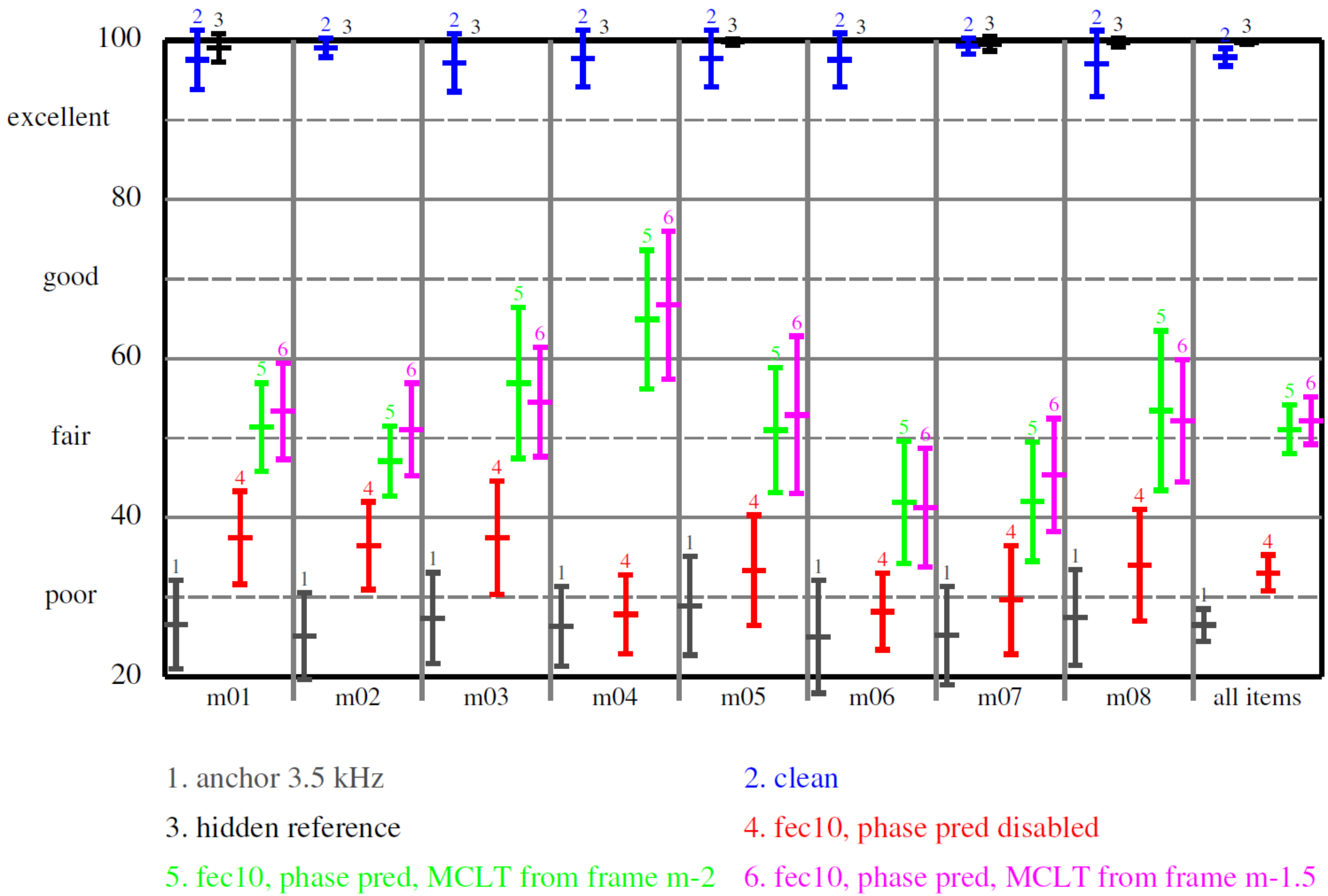
- Find peaks - local maxima l_n



- Spectral envelope obtained using equally weighted moving average filter with pitch dependent adaptive filter order
- The existence of a peak in frames $m-2$ and $m-1$ differentiates true peaks from noise
- Tonal component is the peak and 3 neighboring coefficients on each side

Fraunhofer IIS

Listening test results



Fraunhofer IIS

Conclusions

- Distortions in tonal components lead to perceptually annoying artefacts
- Phase estimation preserves the continuity of tonal components
- Robust tonal component detection was achieved
- EVS constraints were fulfilled

Perfect conditions

Noise substitution

Proposed method

Thank you for your attention!

Questions?

Fraunhofer IIS