# PRE-ECHO REDUCTION IN TRANSFORM AUDIO CODING VIA TEMPORAL ENVELOPE CONTROL WITH MACHINE LEARNING BASED ESTIMATION

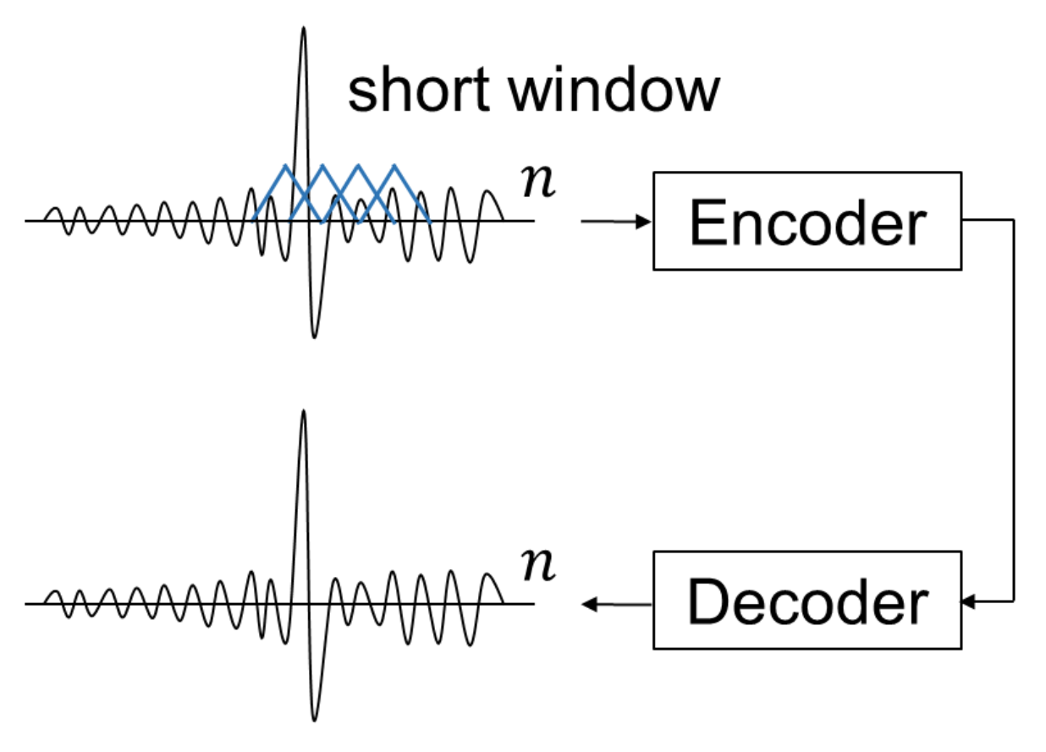
*Jae-Won Kim*[1], *Byeongho Jo*[2], *Seungkwon Beack*[2] *and Hochong Park*[1]

[1]Kwangwoon University, Seoul, Korea

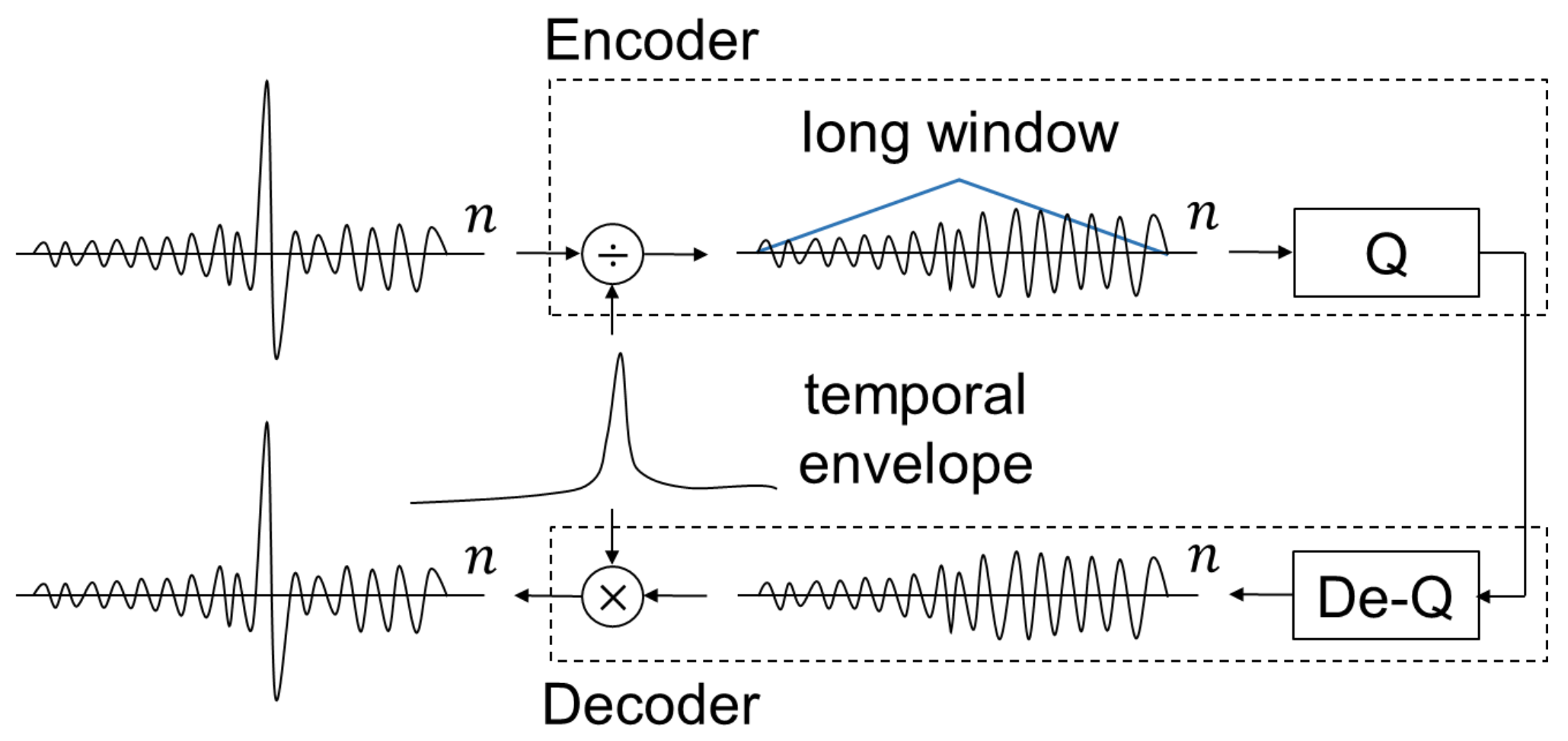
[2]Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea

## Introduction

- New pre-echo reduction method via temporal envelope control with machine learning based estimation
- Novelty of proposed method
  - Direct modification of waveform based on temporal envelope before encoding and after decoding
  - Machine learning based estimation of temporal envelope from side information
  - New coding tool for pre-echo reduction for legacy transform codecs



< Conventional window switching >

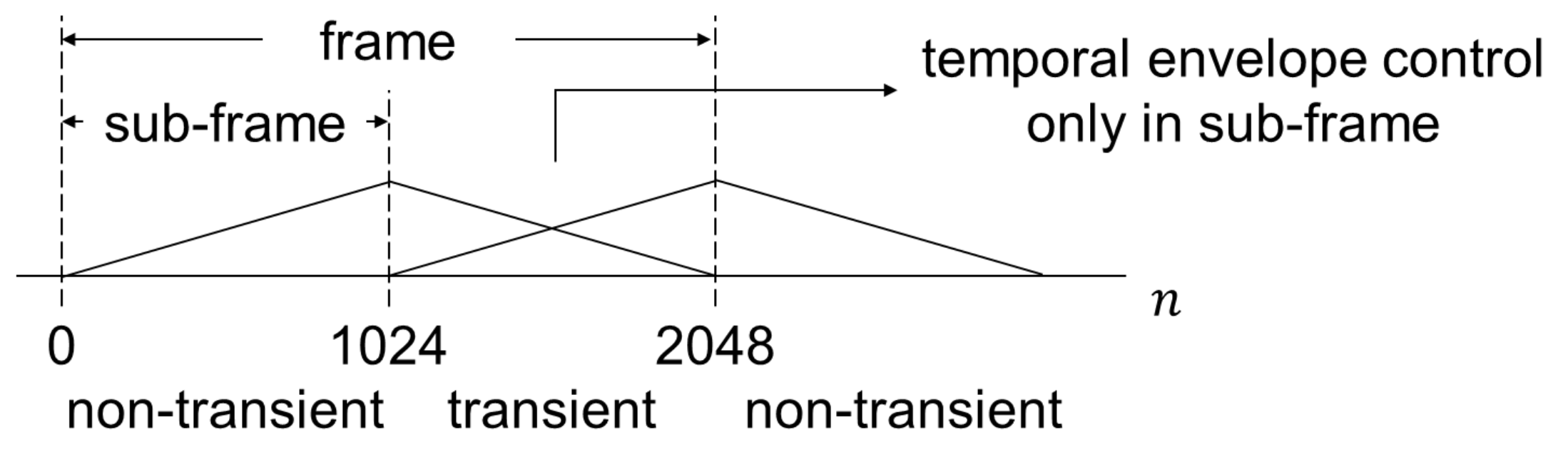


< Proposed method >

- Performance
  - Equivalent sound quality to short-window transform using fewer bits in transient frames

## Proposed method : Two operating modes

- Sub-frame-based temporal envelope control

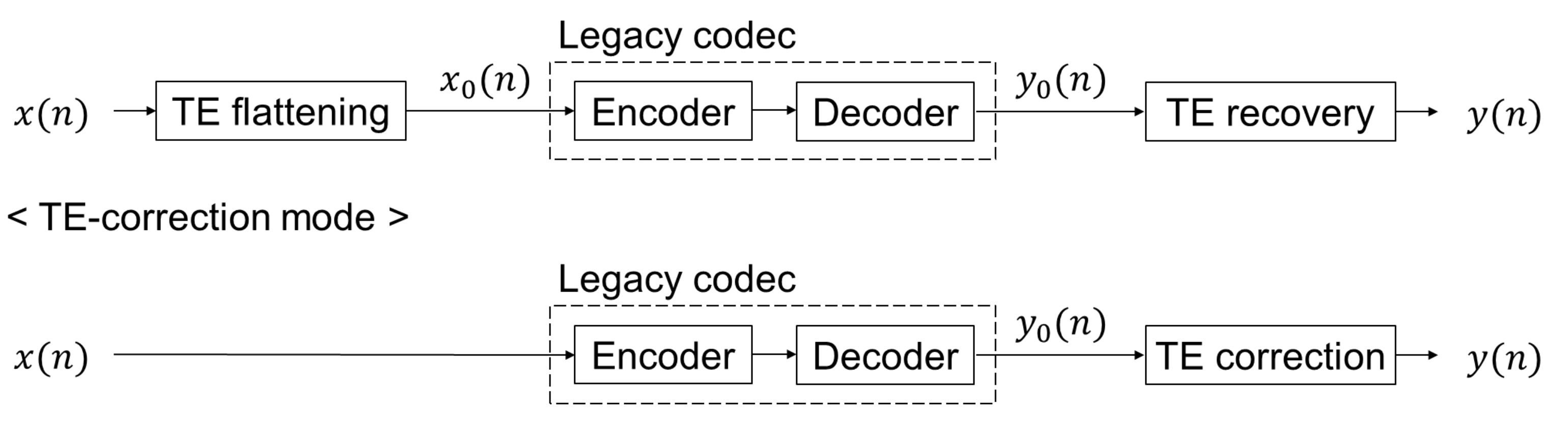


< Sub-frame based temporal envelope control >

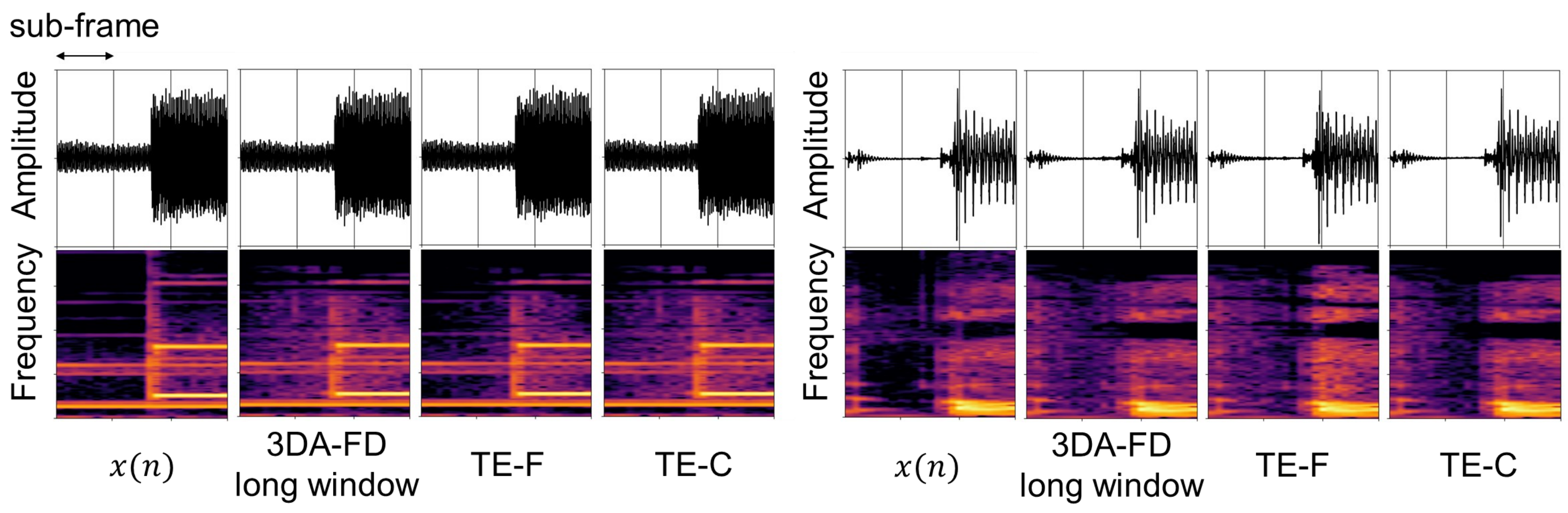
- Two operating modes
  - TE-flattening (TE-F) mode
    - Good pre-echo reduction performance for most transient signals
  - TE-correction (TE-C) mode
    - Effective for some on-set speech signals

< TE-flattening mode >

< TE-correction mode >



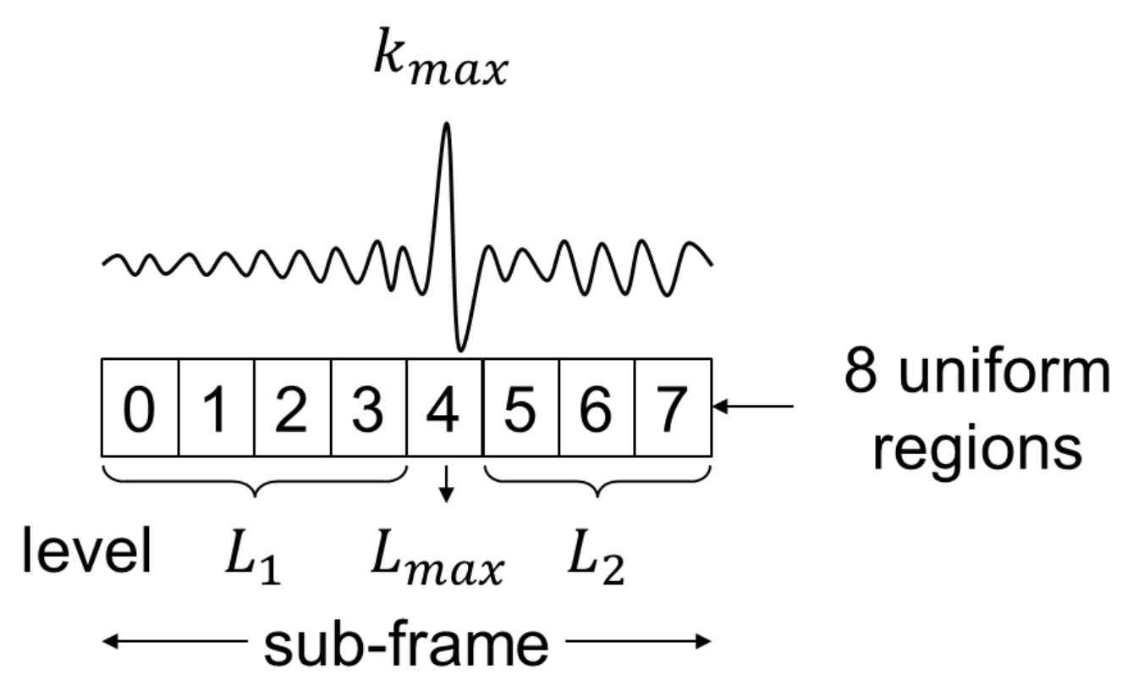
< Two operating modes in the proposed pre-echo reduction method >



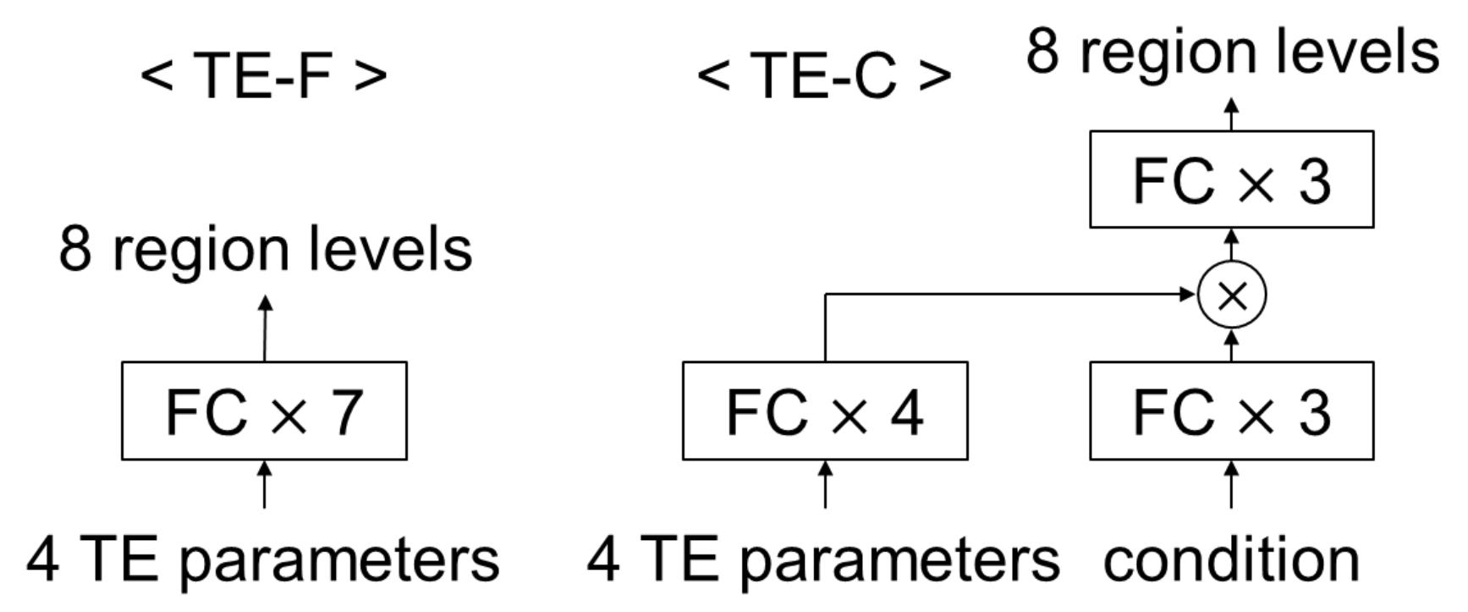
< Output comparison between TE-F and TE-C >

## Proposed method : Envelope prediction

- Temporal envelope estimation using TE parameters and neural network
- 4 TE parameters
  - Max region index $k_{max}$ (3 bits) and max region level $L_{max}$ (3 bits)
  - Level ratio $L_1/L_{max}$, $L_2/L_{max}$ (5 bits each)



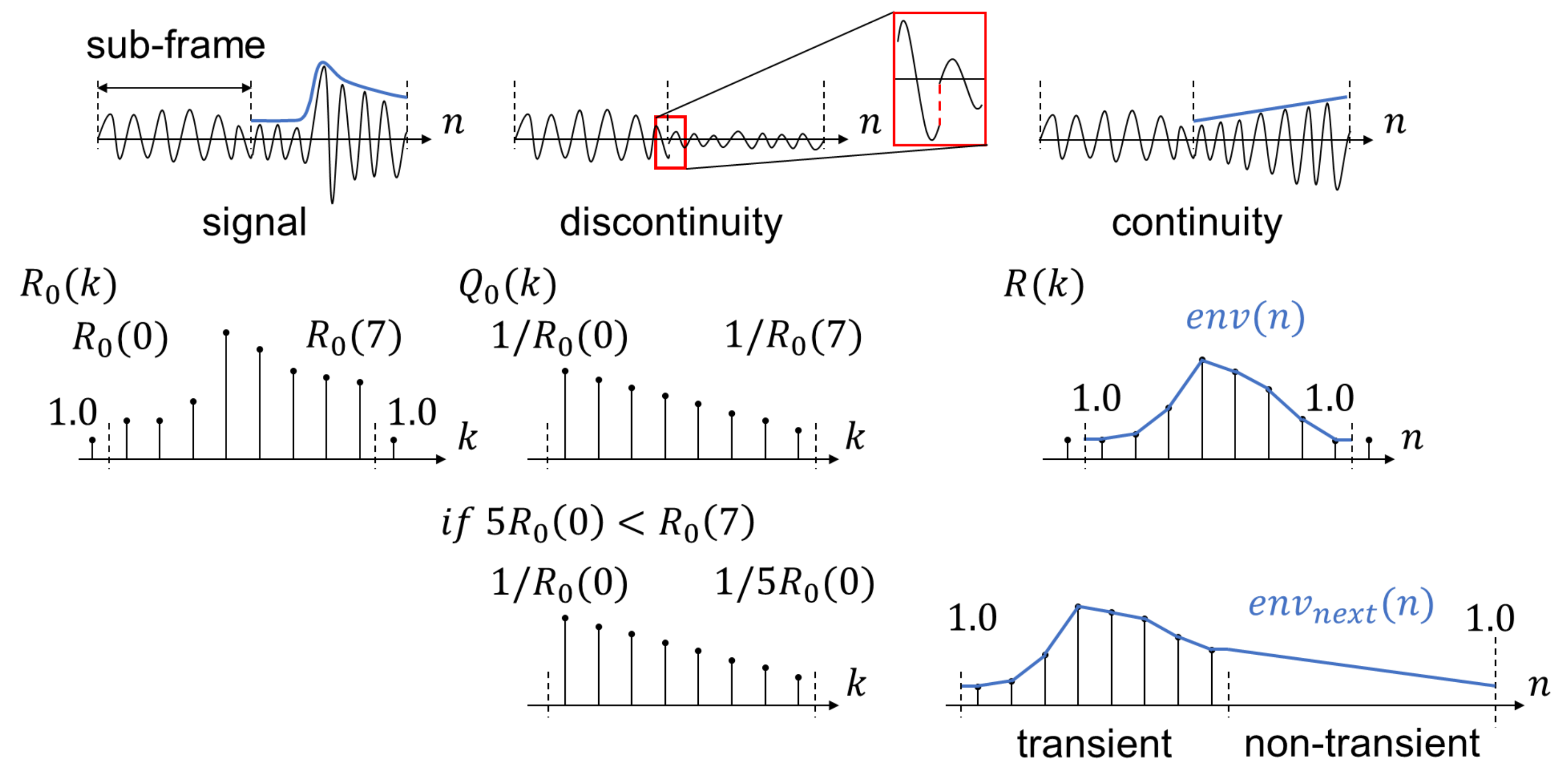
< Extraction of parameters >



< Neural network architecture >

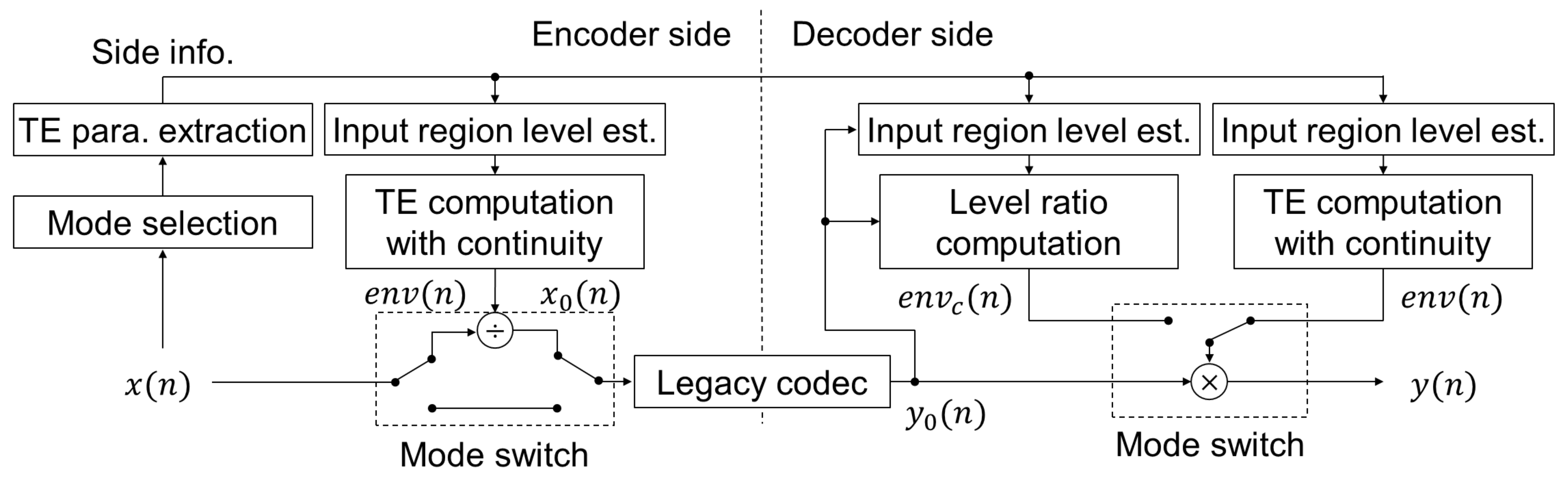
## Proposed method : Continuity

- Temporal envelope computation for frame continuity



< Example of continuity enforcement for TE-F >

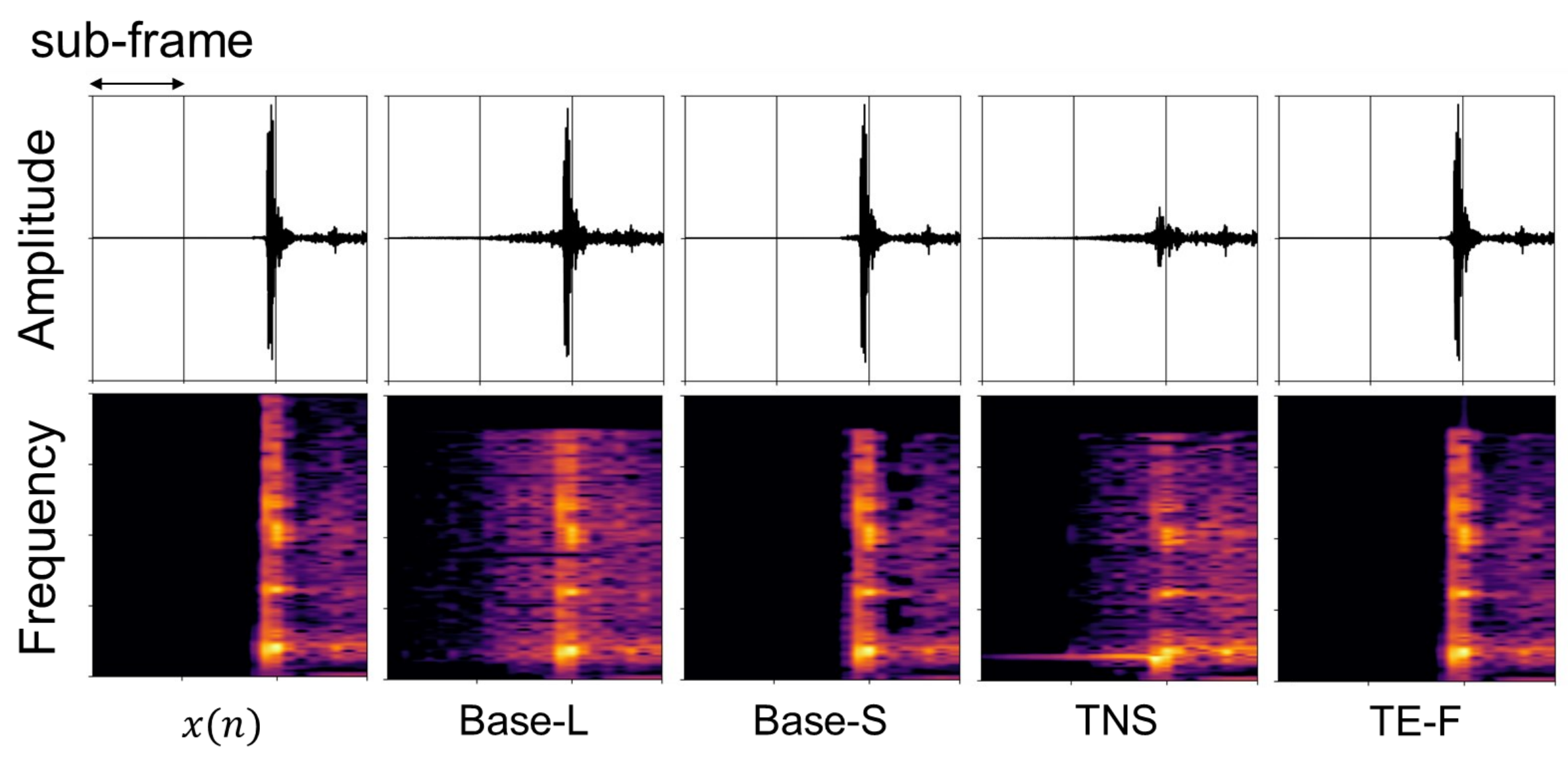
## Proposed method : Overall operation



< Block diagram of overall operation >

## Performance evaluation

- Database
  - Train/validation data : transient signals extracted from Beethoven sonata, VCTK dataset, RWC music database (total 2.5 hours)
  - Test data : 10 audio clips with frequent transient frames (60 sec)
- Core codec : MPEG-H 3D Audio Frequency-Domain mode (3DA-FD)
  - Use transient frames determined by window selection module in the 3DA-FD
- Manually selected operating mode for each transient frame
- Comparison with various pre-echo reduction methods
  - 3DA-FD using long window (Base-L) and short window (Base-S)

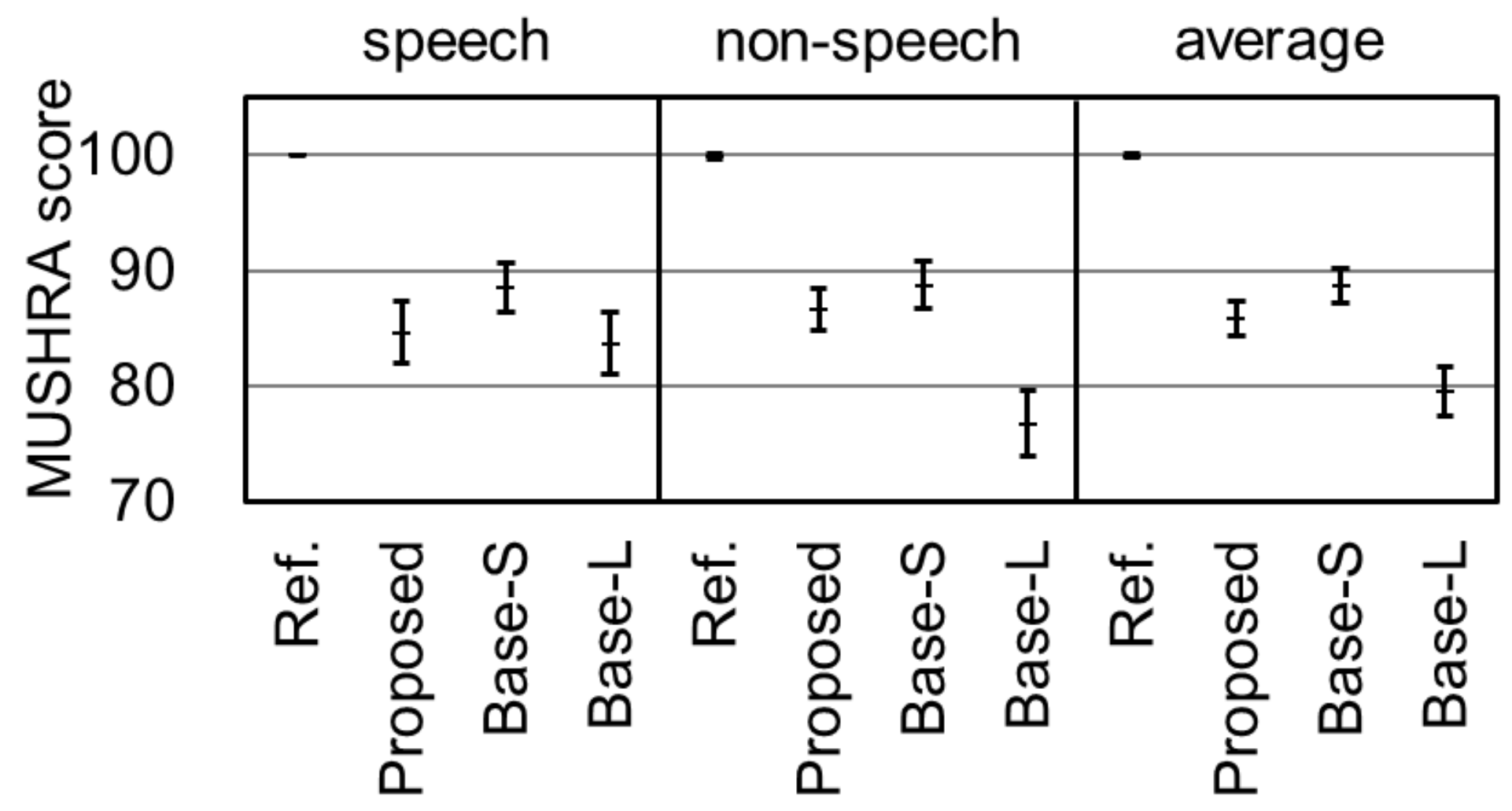


< Results of various pre-echo reduction methods >

- Comparison of average bit rate in transient frames for each method

| Method | Base-L | Base-S | TNS | Proposed |
|---|---|---|---|---|
| Bit rate (kbps) | 47.48 | 49.48 | 48.03 | 47.70 |

- Subjective performance evaluation by MUSHRA
  - Equivalent performance to Base-S using fewer bits
  - Significantly better performance than Base-L using a similar number of bits



## Conclusion

- The proposed method reduces the pre-echo in transform coding by controlling temporal envelope before encoding and after decoding.
- The proposed method using fewer bits yields equivalent sound quality to the short-window transform for mono coding.