# Comparison of Voice Activity Detection Algorithms for Wireless Personal Communications Systems

*Khaled El-Maleh* *Peter Kabal*

McGill University
Department of Electrical Engineering

## Presentation Overview

- The voice activity detection (VAD) problem
- VAD applications
- VAD design
- VAD algorithms
- Comparative study
- Improving VAD performance
- Summary of results

## The Voice Activity Detection Problem

- Conversational (dialogue) speech: sequence of segments of speech and silence
- Background acoustical noise contaminates the speech signal resulting in either speech-plus-noise, or noise-only periods
- An ON-OFF model of conversational speech is given as:

$$x(k) = \begin{cases} s(k) + n(k); & \text{talk mode} \\ n(k); & \text{listen mode} \end{cases}$$

- The VAD problem can take the form of a binary hypotheses testing problem:

- Null hypothesis ($\mathcal{H}_0$): noise-only
- Alternative hypothesis ($\mathcal{H}_1$): speech-plus-noise

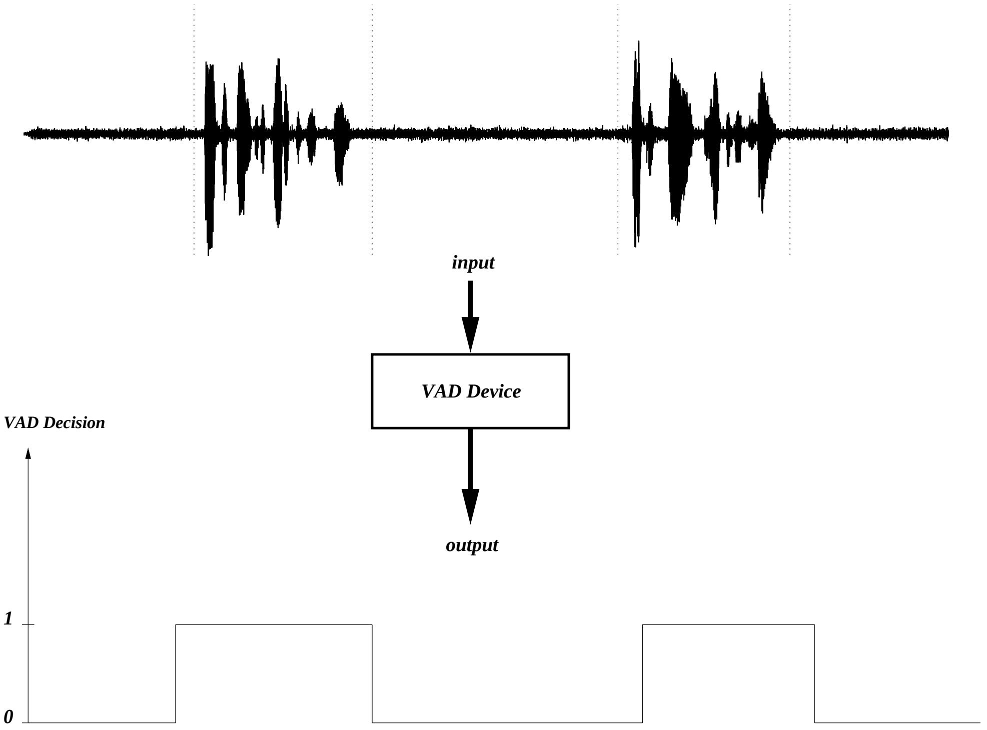

**Fig. 1** The VAD Problem

## VAD Applications

- *Speech Coding*
  - Variable bit rate coding (i.e. QCELP, EVRC)
  - Discontinuous transmission (i.e. GSM coders, G.723.1)
  - Digital speech interpolation (DSI)
- *Speech Recognition*
- *Echo cancellation* (hands-free telephony, audio-conferencing)
- *Noise reduction systems* (i.e. spectral subtraction algorithms)
- *Speech synthesis*

# Basics of a VAD Design

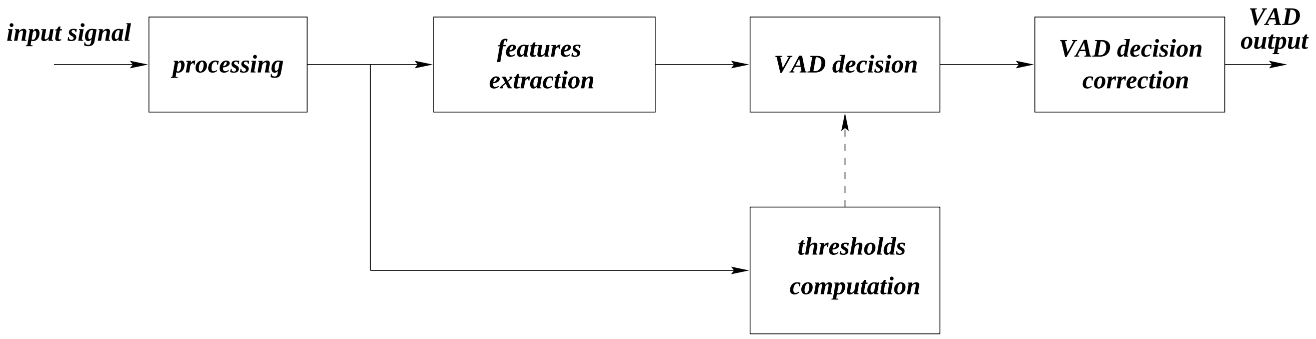

**Fig. 2** A General VAD Algorithm

*Commonly used VAD features:*

- short-time energy
- zero crossing rate
- LPC, and cepstral coefficients
- Pitch lag (periodicity)

# The GSM VAD

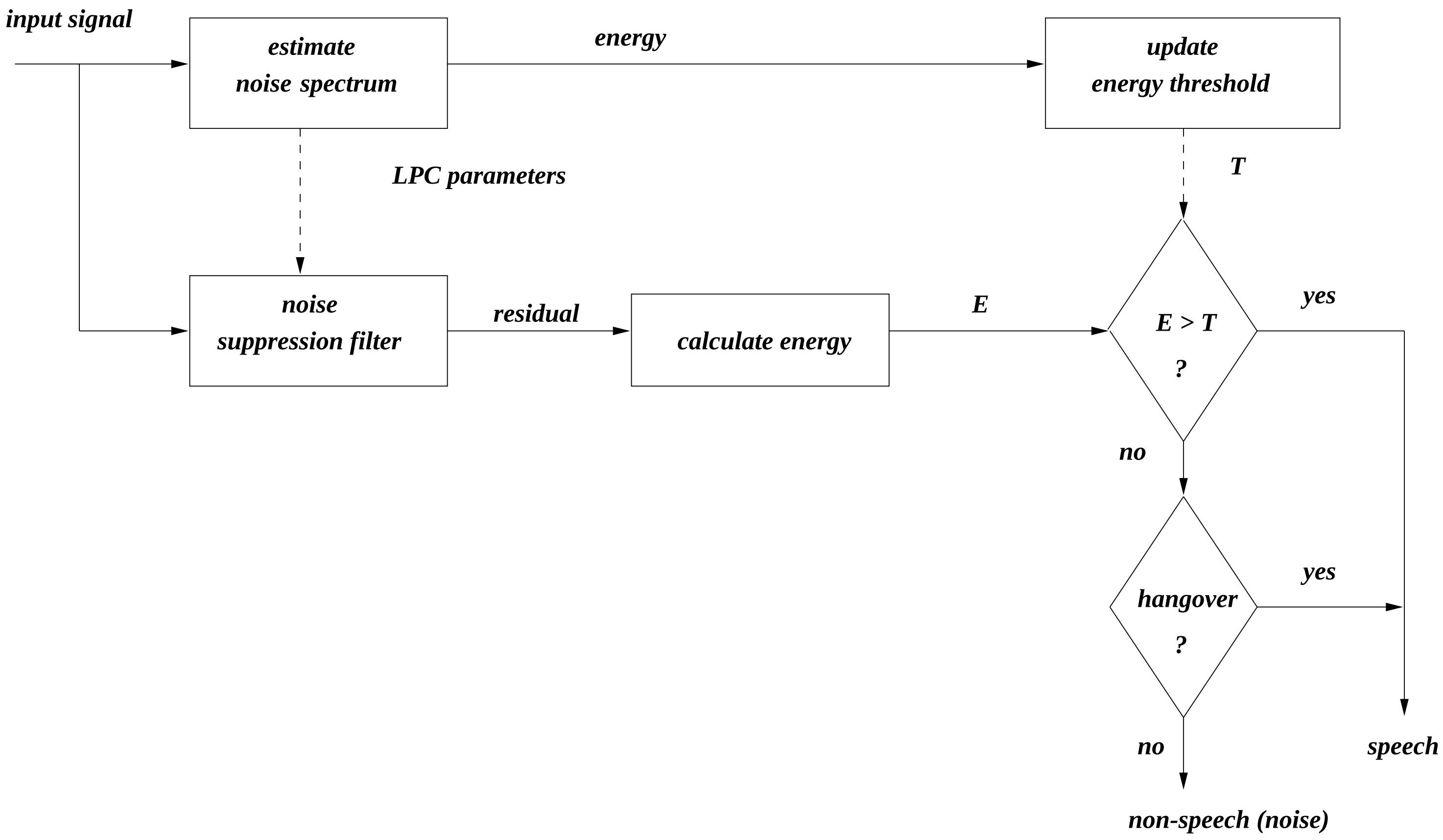

**Fig. 3** The GSM VAD Algorithm

## The Improved GSM VAD

- Srinivasan and Gersho (1993) proposed an improved version of the GSM VAD
- Several new features to the basic GSM VAD design include:
  - a multi-band (4 bands) energy comparison
  - spectral flatness measurements
  - using the fraction of the energy of the low frequency band

# The EVRC VAD

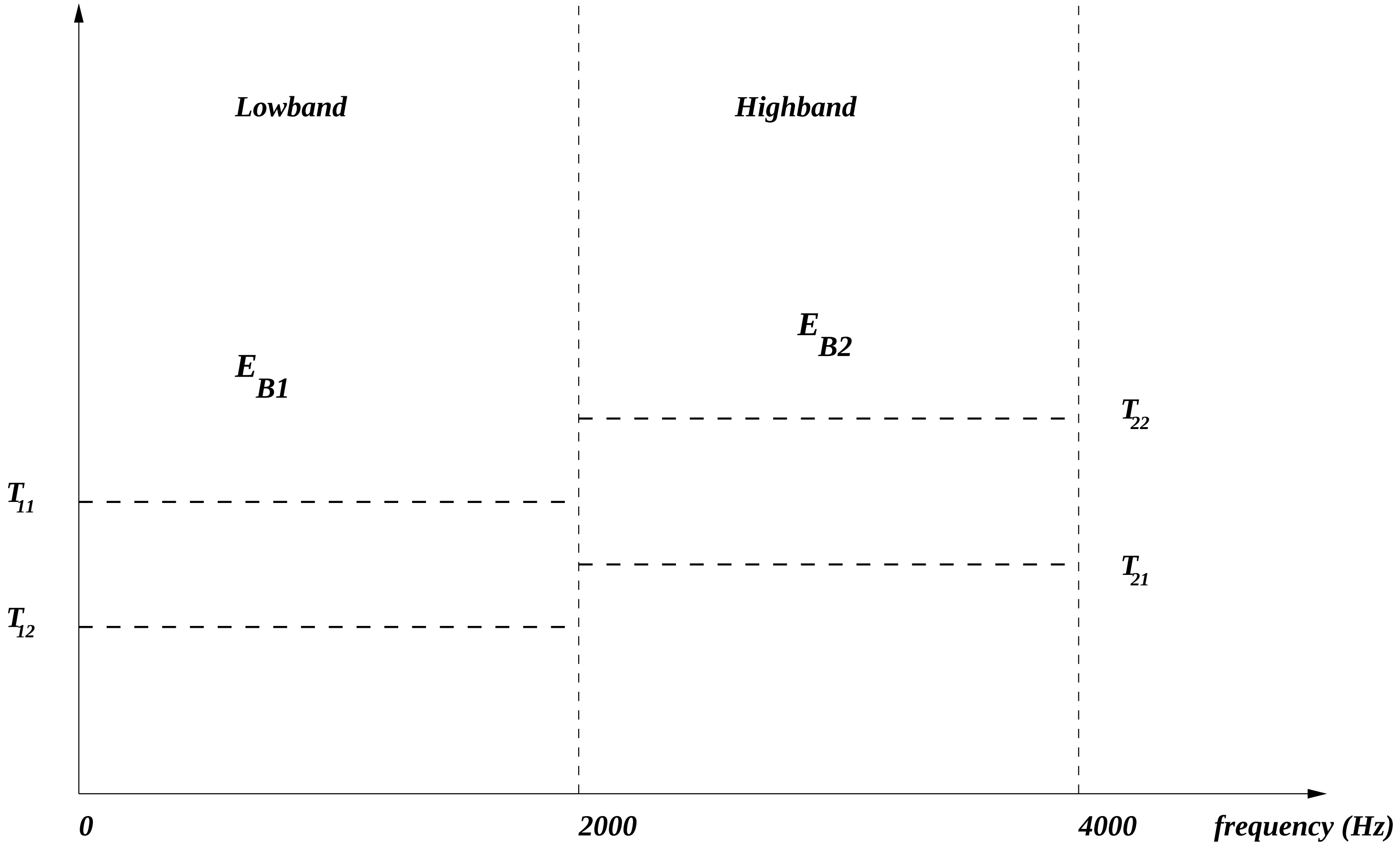

**Fig. 4** EVRC VAD Thresholding Mechanism

# The Third-Order Statistics VAD

- Symmetrically distributed (non-skewed) processes have a zero *third-order cumulant* (TOC) at all lags

- Speech: skew enough to have significantly non-zero TOC at all lags

- Many real-life noises can be assumed to be Gaussian or at least symmetrically distributed

- A time domain Gaussianity test is used as the basis for the *third-order statistics* (TOS) VAD

- The test statistic of this VAD is defined as

$$\hat{d} = \hat{c}_{3y}^{t} \hat{C}_0^{-1} \hat{c}_{3y}$$

  - $\hat{c}_{3y}$: TOC of a given frame
  - $\hat{C}_0$: covariance matrix of the TOC estimated from $R$ initial noise-only frames

- VAD threshold ($\mathcal{T}$): $\chi_Q^2(\alpha)$

  - $\alpha$ is a pre-selected probability of false alarm ($P_F$)
  - $Q$ is the number of lags used in the TOC computation

- The value of the threshold is obtained from the chi-square ($\chi_Q^2$) table

# VAD Hangover Algorithms

- In VAD algorithms, a hangover (HOV) period of few frames (3–6) are used to prevent premature transition from speech to noise
- HOV algorithms are used to avoid detecting low-energy unvoiced speech as noise
- Both the GSM and the EVRC VADs use HOV algorithms
- EVRC VAD uses an adaptive hangover period based on the SNR estimate of each frame

## Comparative Study

- Compare the performance of each VAD under different acoustical background noise conditions
- These conditions include different noise environments (street, car, bus, and restaurant) and at various signal-to-noise (SNR) ratios (20, 10, and 0 dB)
- Two types of VAD errors:
  - detection of speech as noise (probability of miss)
  - detection of noise as speech (probability of false alarm)

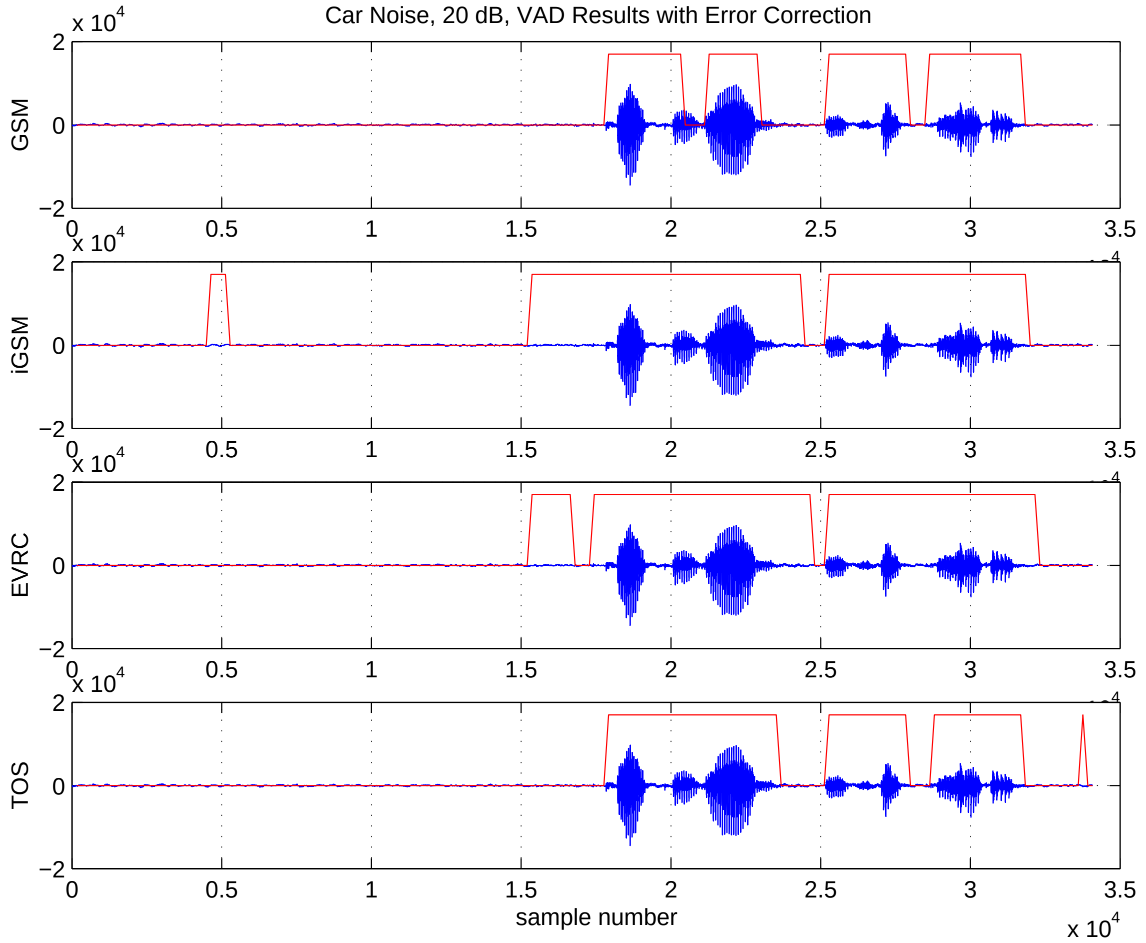

**Fig. 5** VAD results for car noise at 20 dB SNR.

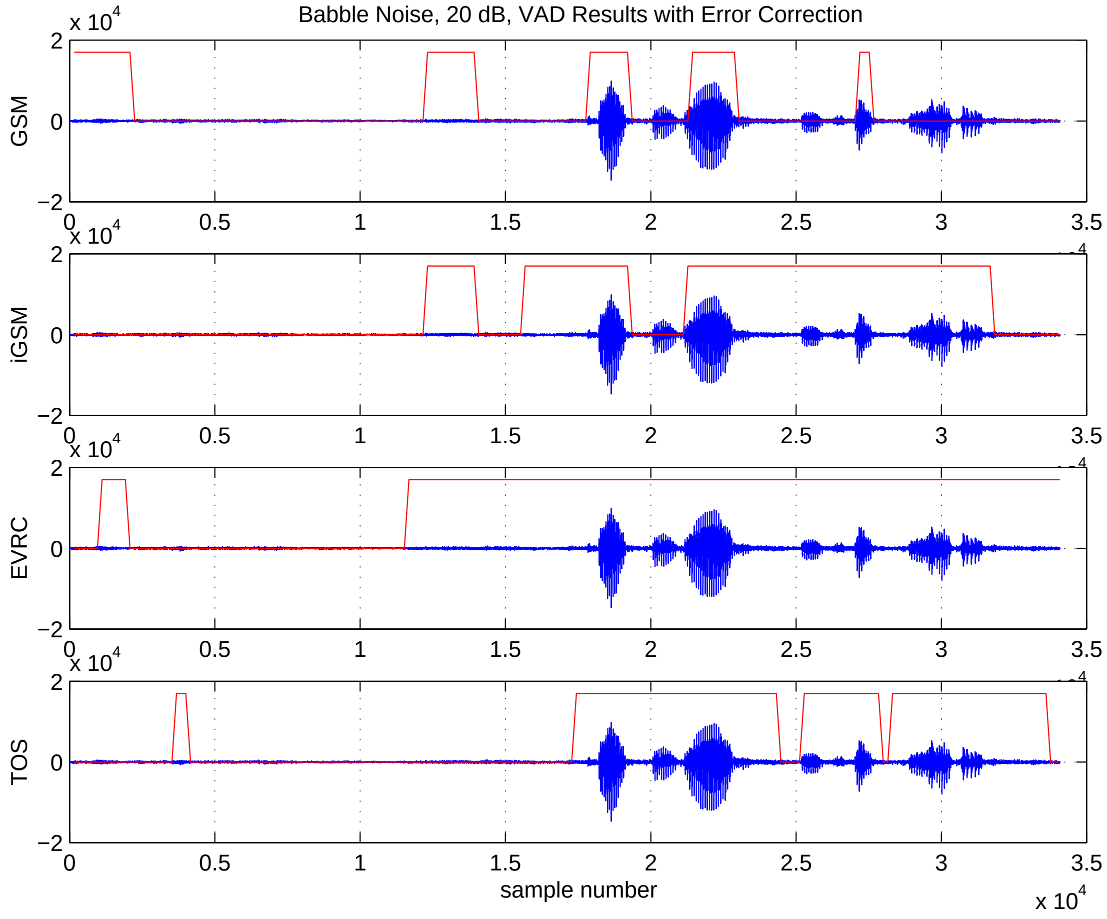

**Fig. 6** VAD results for babble noise at 20 dB SNR.

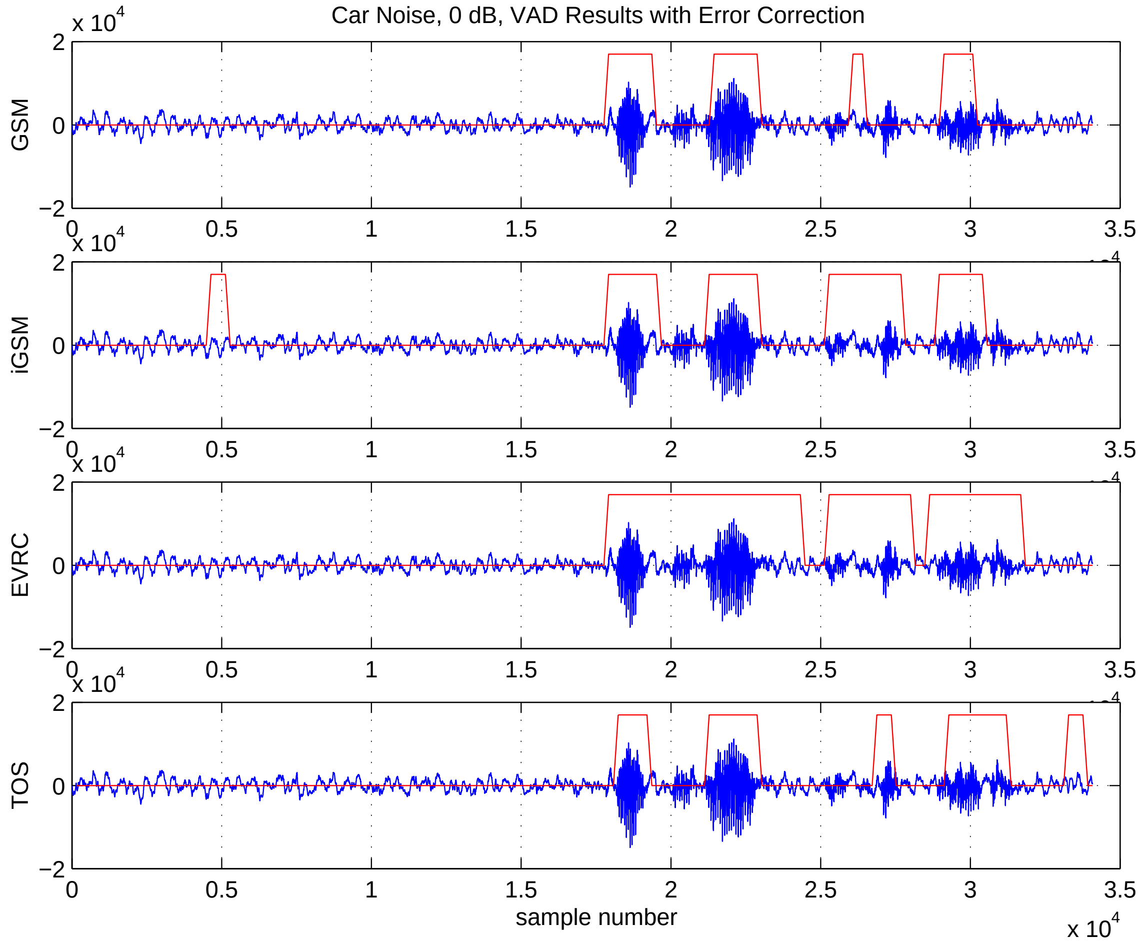

**Fig. 7** VAD results for car noise at 0 dB SNR.

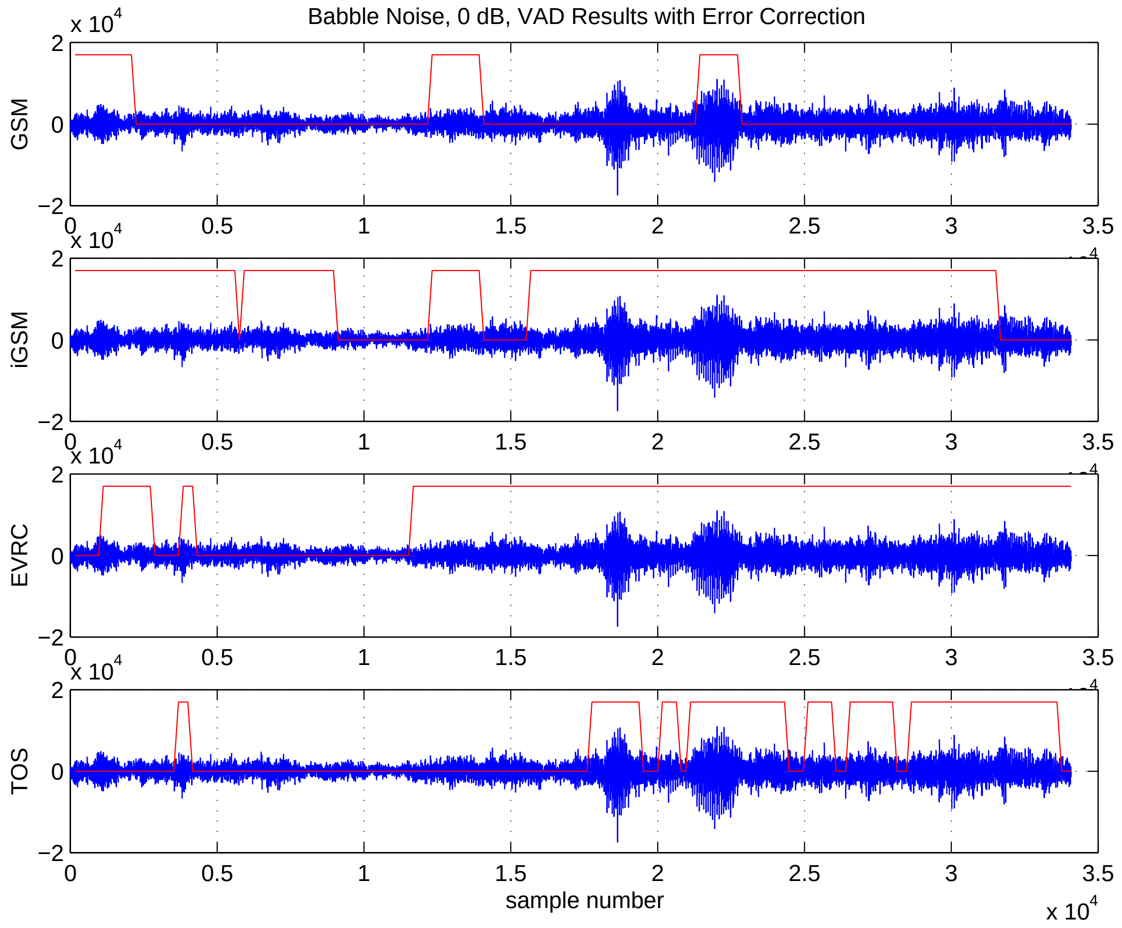

**Fig. 8** VAD results for babble noise at 0 dB SNR.

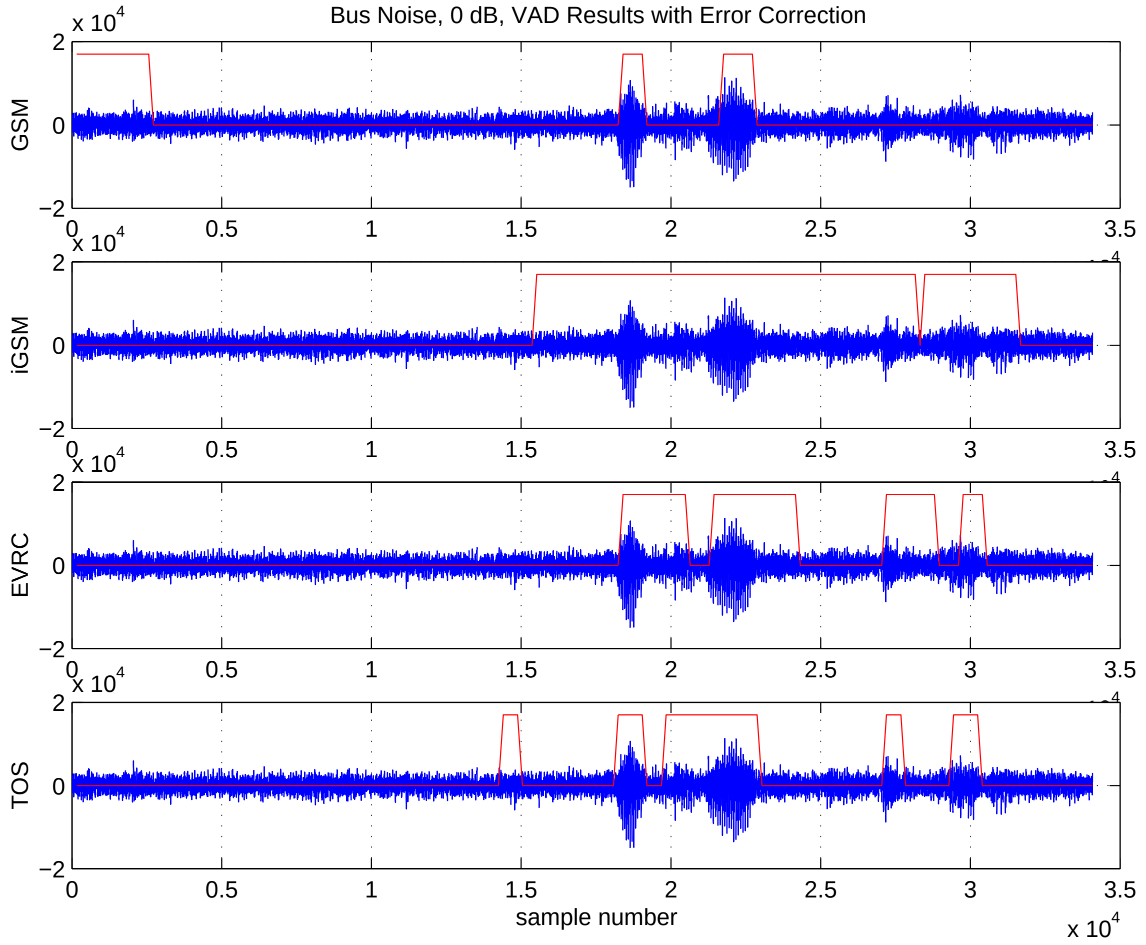

**Fig. 9** VAD results for bus noise at 0 dB SNR.

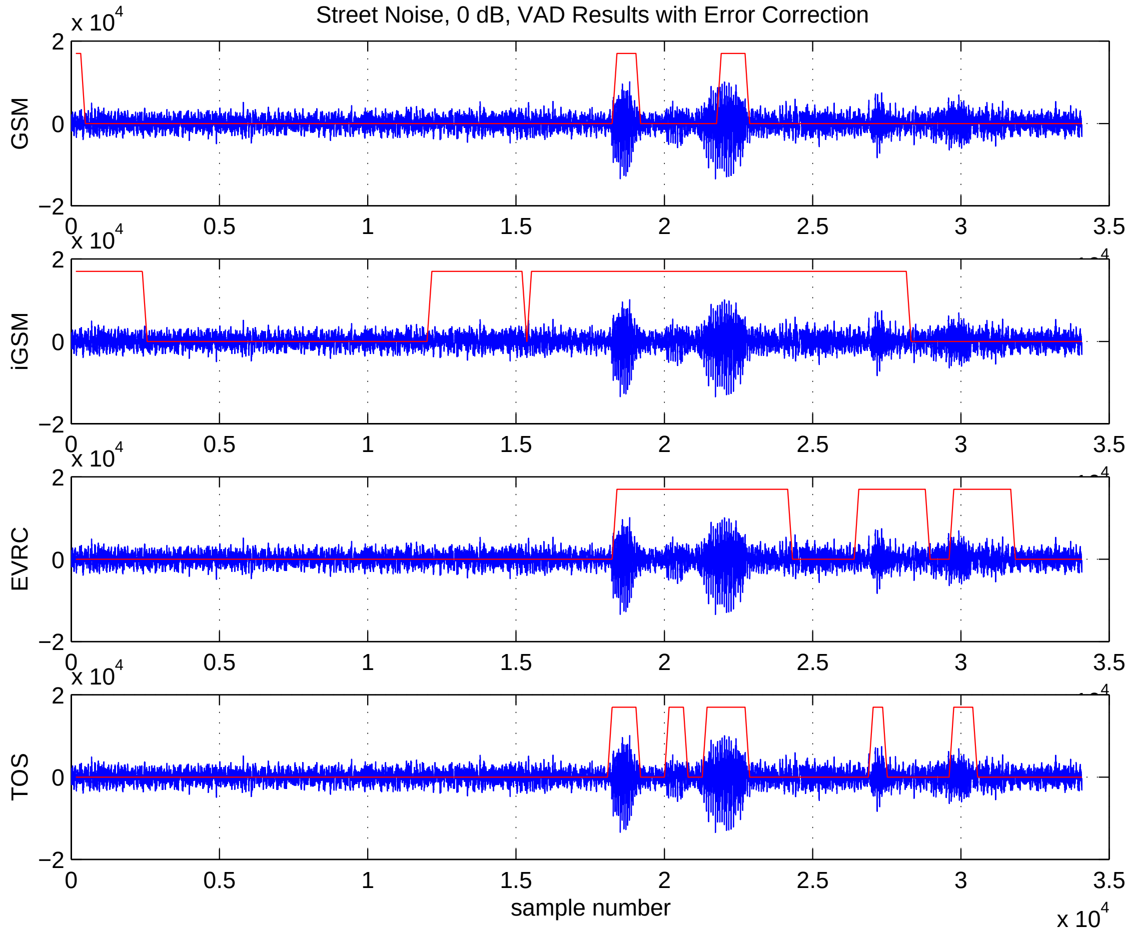

**Fig. 10** VAD results for street noise at 0 dB SNR.

# Improving VAD Performance

- Use the linear prediction (LP) residual as the input signal to the VAD
- Isolated VAD errors result in annoying perceptual artifacts in VBR speech coders
- Isolated error correction mechanism (IECM)
  - Delay the decision by 2–3 frames to monitor the VAD decisions in neighboring frames
  - If the VAD decision of the current frame is different from its neighbors, then its VAD flag is changed to be similar to the other frames

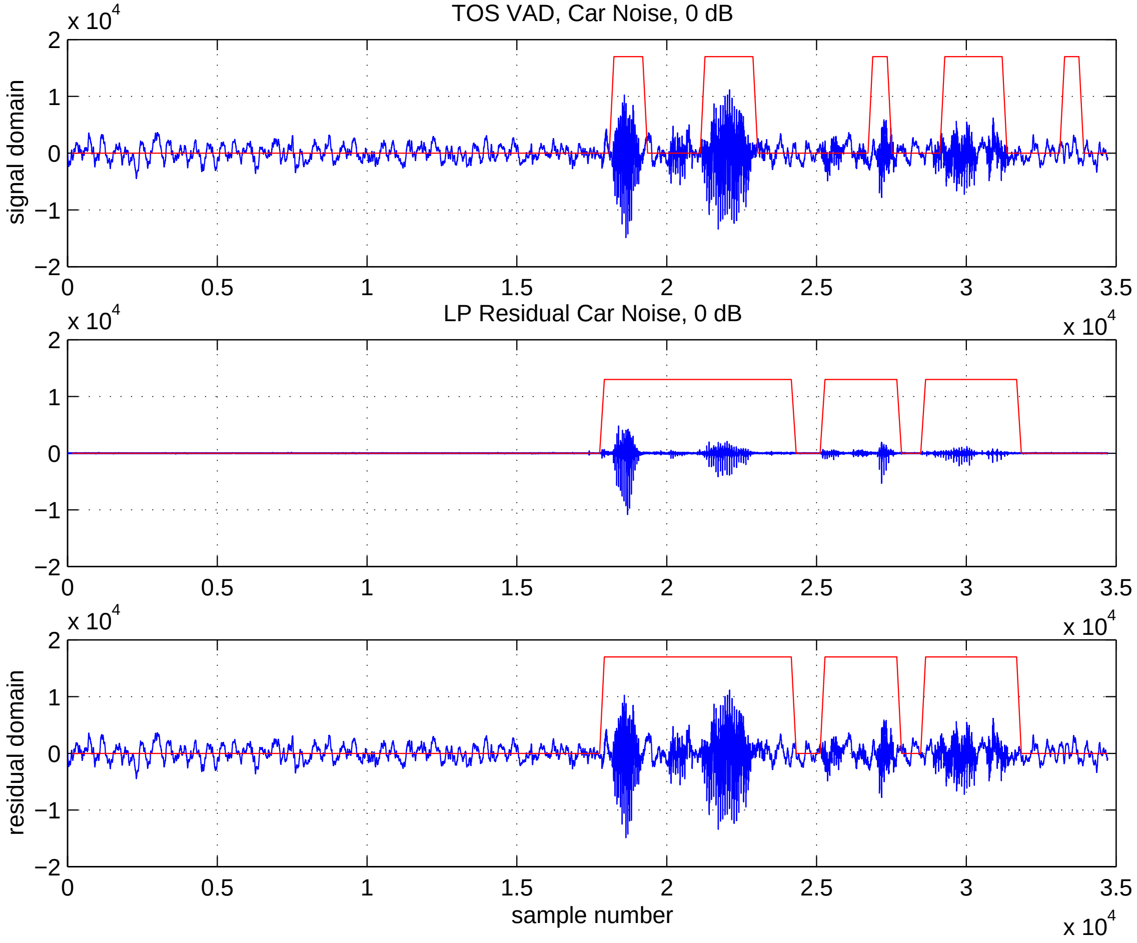

**Fig. 11** TOS VAD: effect of input signal.

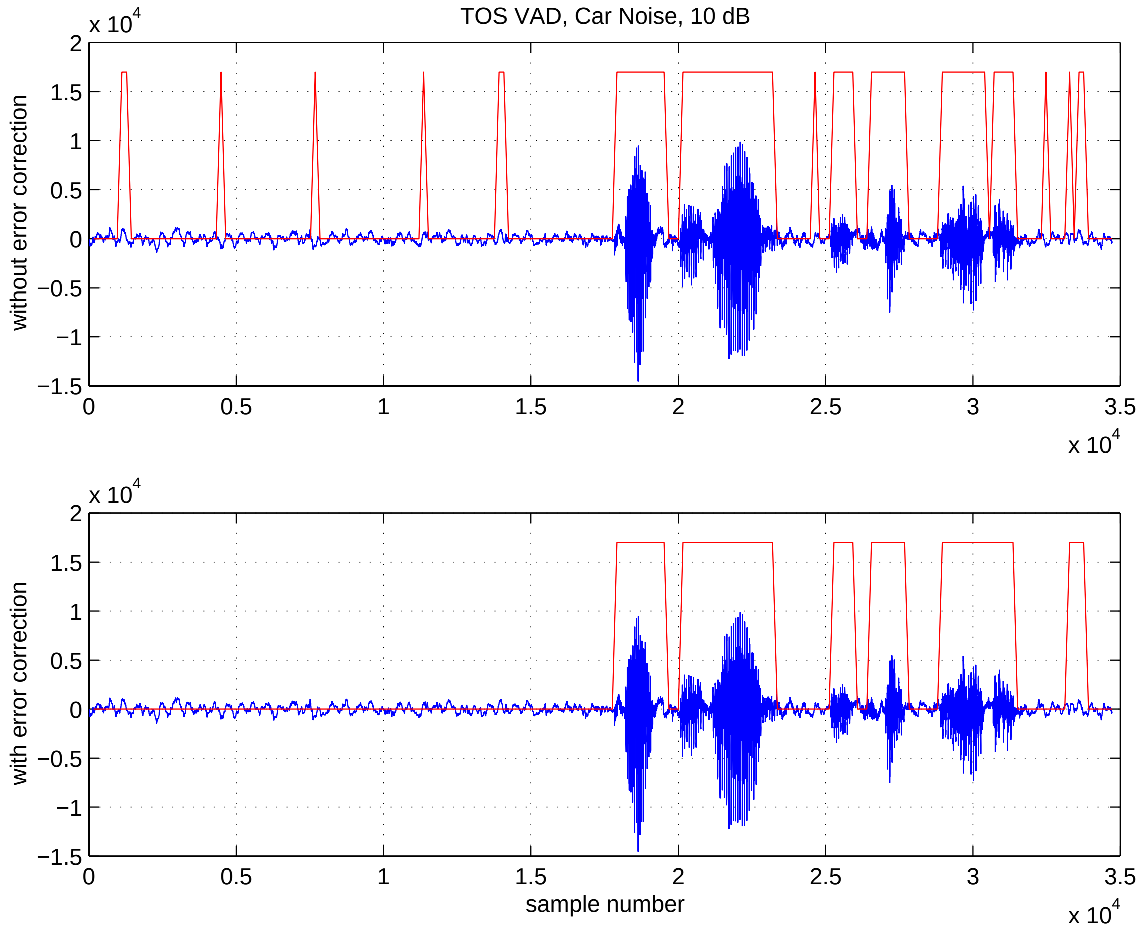

**Fig. 12** TOS VAD: effect of isolated error correction mechanism.

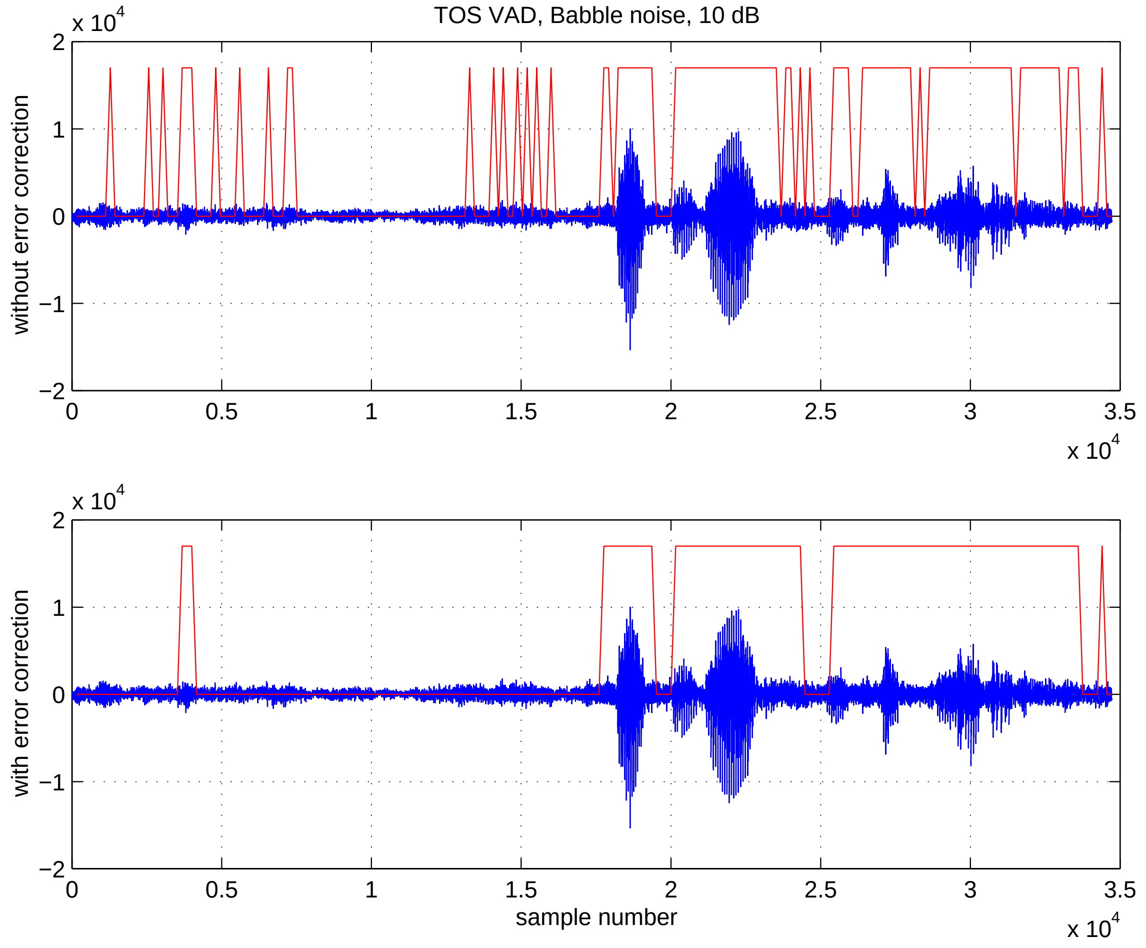

**Fig. 13** TOS VAD: effect of isolated error correction mechanism.

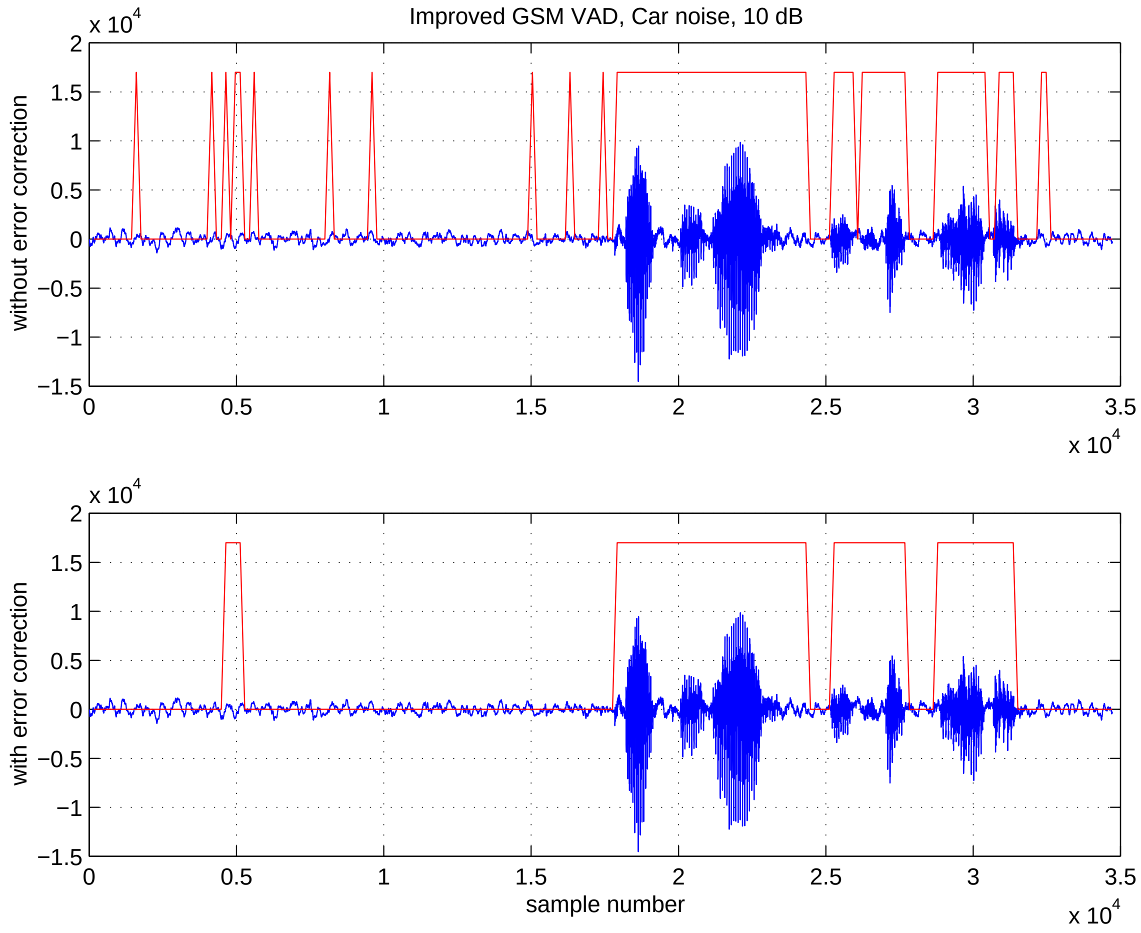

**Fig. 14** TOS VAD: effect of isolated error correction mechanism.

## Summary of Results

- Consistent superiority of the EVRC VAD
- The TOS VAD is ranked overall second in performance with almost-perfect detection of babble noise at 0 dB
- The GSM VAD shows acceptable performance under stationary noise environments but is not good for non-stationary noises
- High-energy voiced speech segments are always detected but low-energy unvoiced speech is commonly missed
- The VAD decisions were improved by using the LP residual as the input signal to the VAD and by using the proposed IECM.

## References

[1] K. Srinivasan and A. Gersho, "Voice activity detection for cellular networks," in *Proc. of the IEEE Speech Coding Workshop.*, pp. 85–86, October 1993.

[2] D. K. Freeman, G. Cosier, C.B. Southcott, and I. Boyd, "The voice activity detector for the Pan-European digital cellular mobile telephone service," in *Proc. Intl. Conf. Acoust., Sp., & Sig. Proc.*, pp. 369–372, Glasgow, May 1989.

[3] TIA Document, PN-3292, Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems, January, 1996.

[4] L. R. Rabiner, and M. R. Sambur, "Voiced-unvoiced-silence detection using the Itakura LPC distance measure," in *Proc. Intl. Conf. Acoust., Sp., & Sig. Proc.*, pp. 323–326, May 1977.

[5] J. A. Haigh, and J. S. Mason, "Robust voice activity detection using cepstral features," in *IEEE TENCON*, pp. 321–324, China, 1993.

[6] J. D. Hoyt, and H. Wechsler, "Detection of human speech in structured noise," in *Proc. Intl. Conf. Acoust., Sp., & Sig. Proc.*, pp. II-237–II-240, Australia, May 1994.

[7] R. Tucker, "Voice activity detection using a periodicity measure," in *IEE Proceedings-I*, Vol. 139, No. 4, pp. 377–380, August 1992.

[8] M. Rangoussi, and G. Carayannis, "Higher order statistics based Gaussianity test applied to on-line speech processing ," in *Proc. of the IEEE Asilomar Conf.*, pp. 303–307, 1995.

[9] H. J. M. Steeneken, and F. W. M. Geurtsen, "Description of the RSG.10 noise database," Report IZF 1988-3, TNO Institute for Perception, Soesterberg, The Netherlands.