Perceptual features for automatic speech recognition in noisy environments

Serajul Haque , Roberto Togneri, Anthony Zaknich

Reporter:邱聖權
Professor:陳嘉平

Introduction

- This paper implements two perceptual properties of the peripheral auditory system, synaptic adaptation and two-tone suppression based on ZCPA feature extraction.
- This model is implemented in time domain.
- The temporal-place representation of auditory processing is much less affected by background noise than the rate-place representation.

The ZCPA model

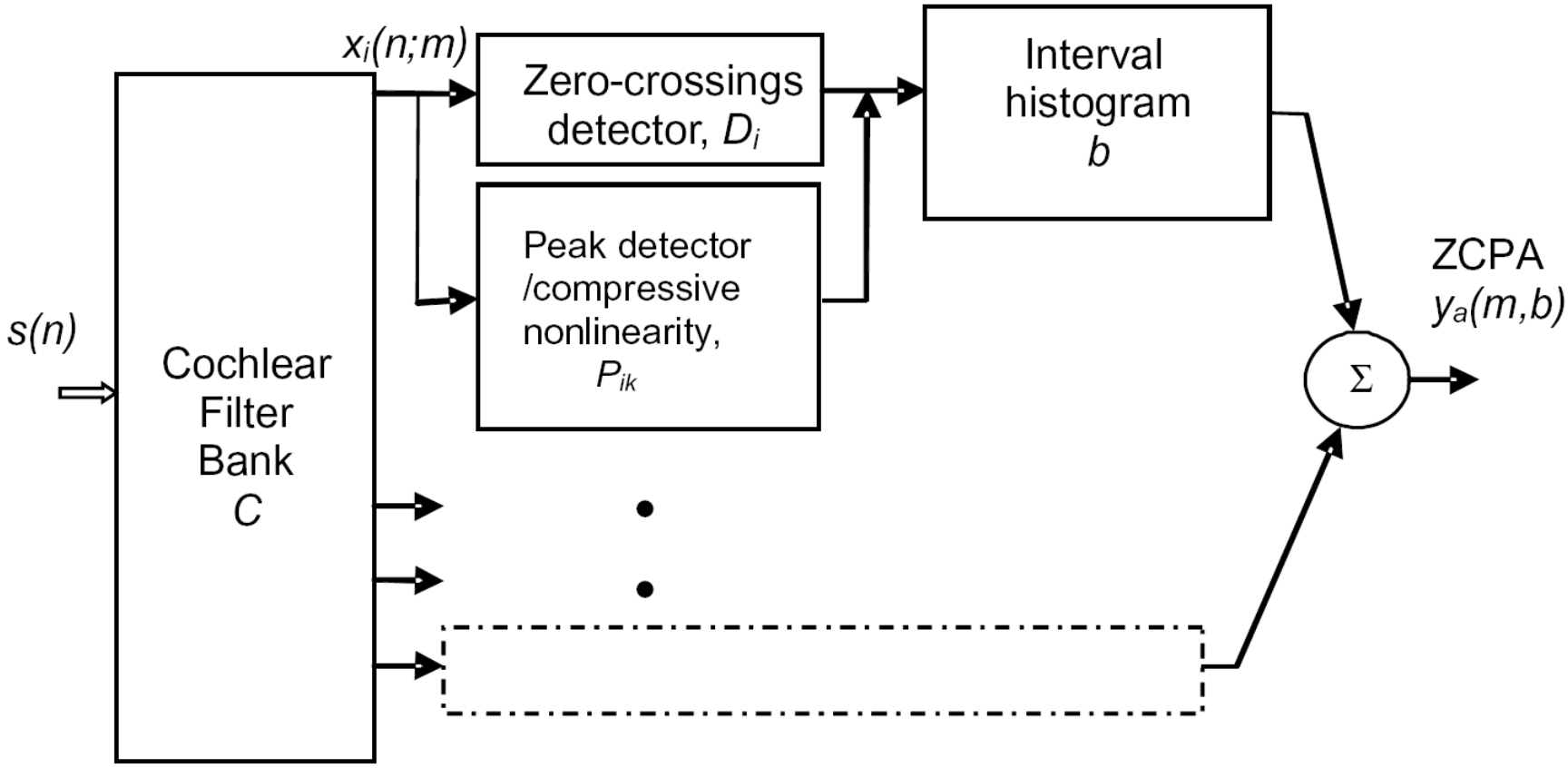


Fig. 2. Schematic of the ZCPA auditory model.

Synaptic adaptation

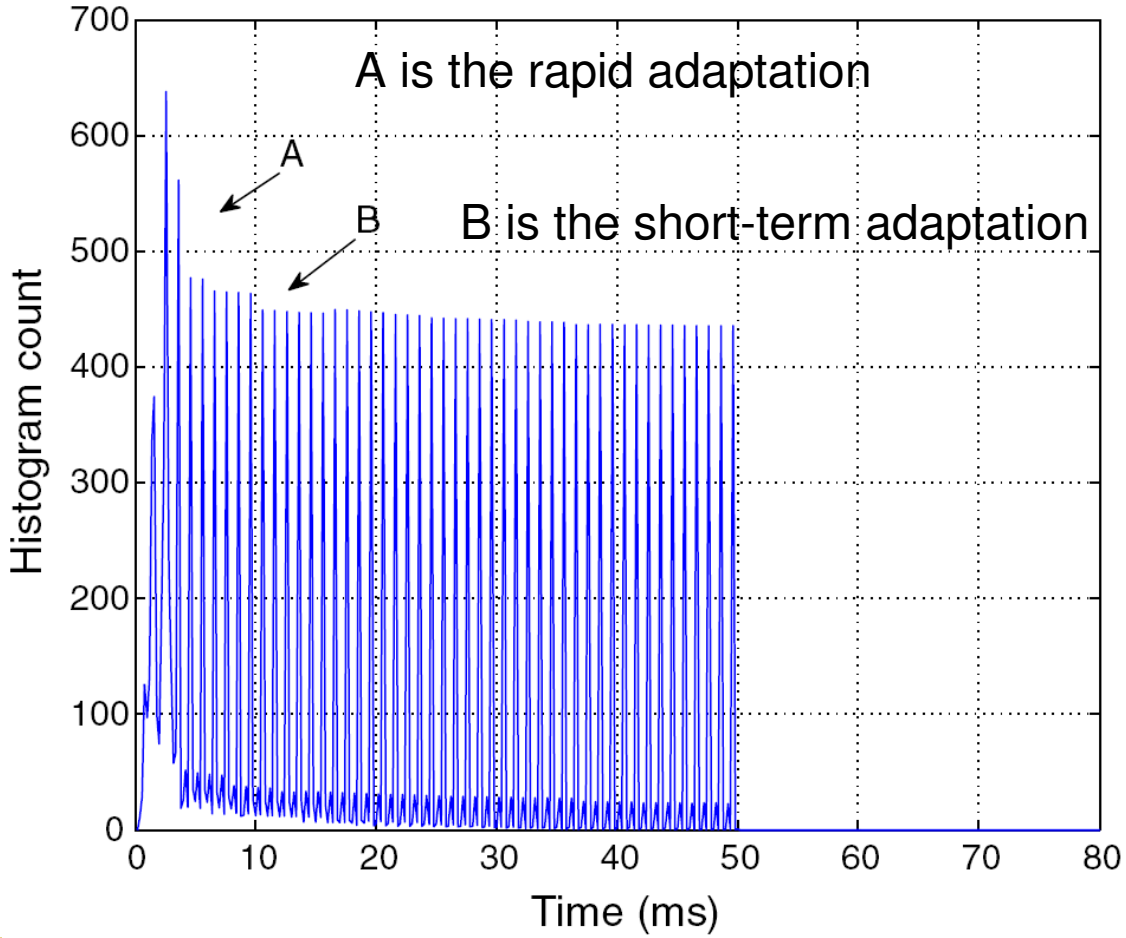
- IIR high pass filter

$$H(z) = \frac{10\tau f_r(1 - z^{-1})}{(10\tau f_r + 0.05) + (10\tau f_r - 0.05)z^{-1}}$$

τ is the time constant setting to 250 ms,

f_r is the frame rate which is 100 Hz

Synaptic adaptation



ZCPA + synaptic adaptation

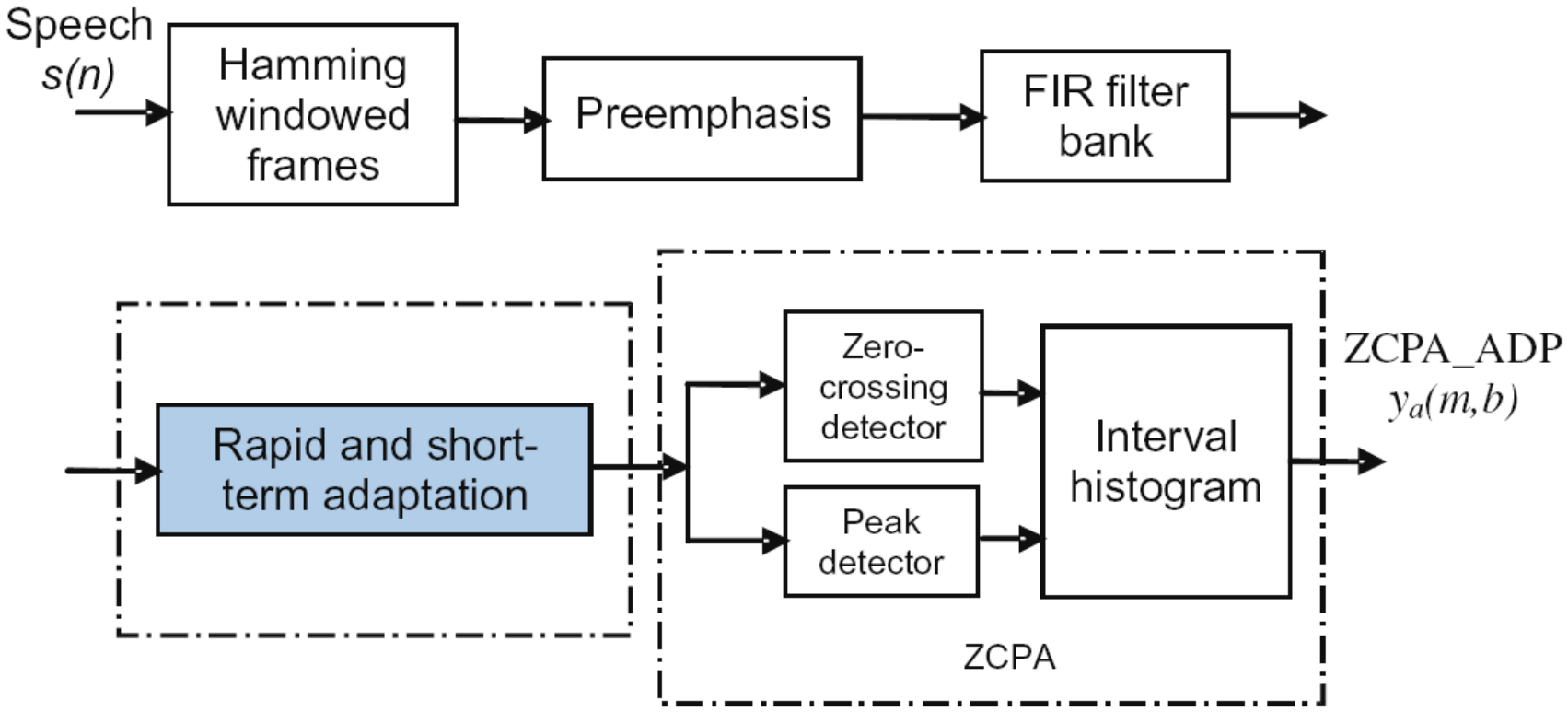
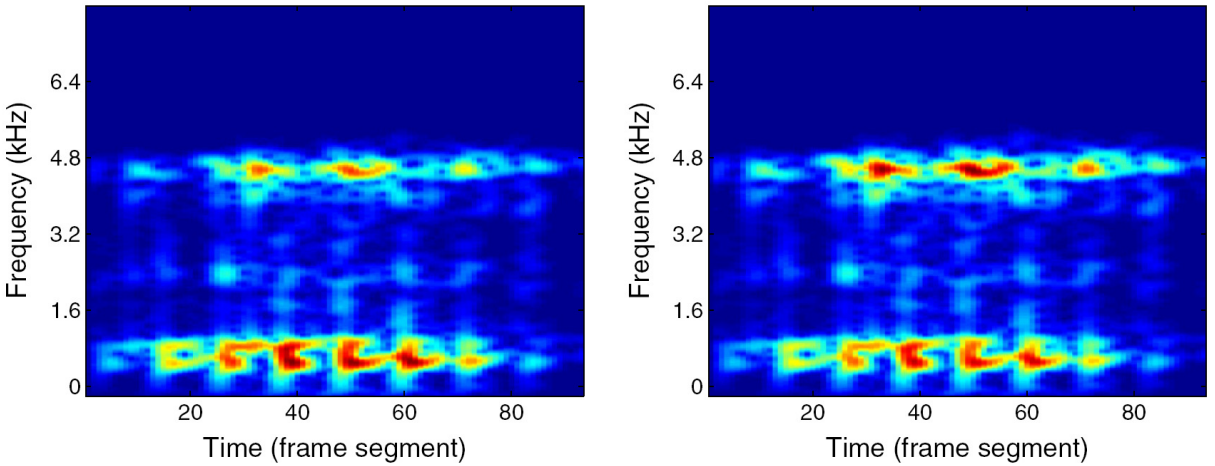


Fig. 5. Schematic of the ZCPA with synaptic adaptation (ZCPA_ADP).

Spectrogram of ZCPA + synaptic adaptation



(a) Without synaptic adaptation /ba/ (b) With synaptic adaptation /ba/

Fig. 6. Spectrogram of the ZCPA for the 35th frame of the male utterance /ba/ in clean condition (a) without synaptic adaptation, and (b) with synaptic adaptation with a time constant $\tau = 250$ ms showing enhanced high-frequency segments.

Two-tone suppression

- Two-tone suppression is the reduction to one tone due to the presence of another tone at a nearby frequency and is a nonlinear property of the cochlea.
- Two-tone can be simulated by companding (compressing and expanding) strategy.

Companding strategy

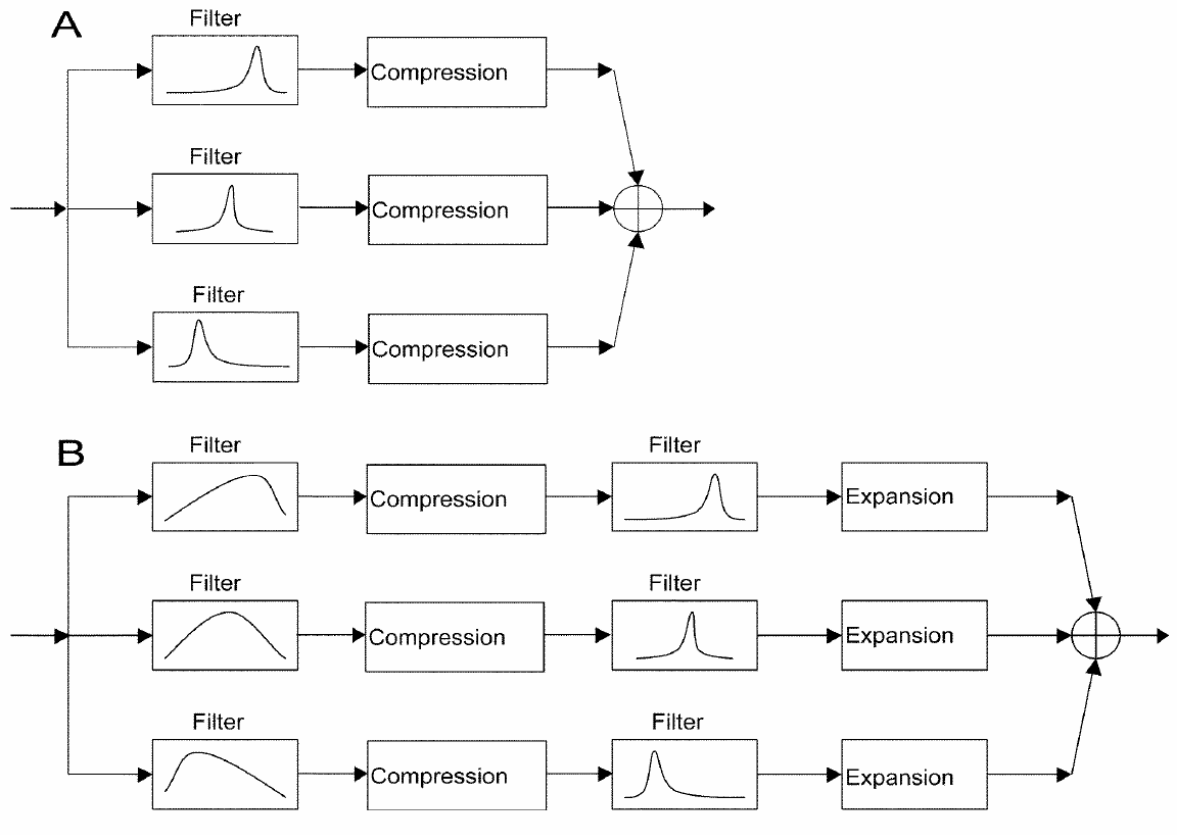


Fig. 1. (a) Block diagram of a common multichannel syllabic compression strategy. (b) Block diagram of our companding strategy.

Multichannel compression

- Multichannel compression by itself improves audibility but degrades spectral contrast
- A weak tone at one frequency is strongly amplified so that it is concurrently audible with a weakly amplified strong tone at another frequency
- The asymmetric amplification due to compression degrades the spectral contrast that was present in the uncompressed stimulus

Multichannel Companding

- By companding, compression is prevented from degrading spectral contrast in regions close to a strong spectral peak while allowing the benefits of improved audibility in regions distant from the peak.

Companding strategy

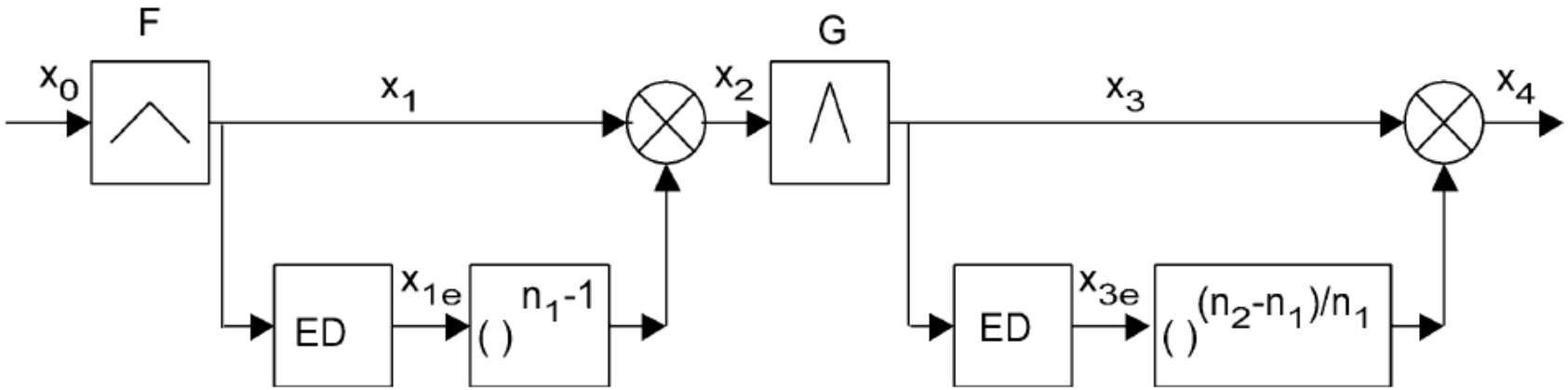


Fig. 2. Detailed view of a single channel of processing in Fig. 1.

ED is a envelope detector, which consisted of a half - wave rectifier followed by a first order low - pass filter.

when $n_1 < 1$, x_0 is compressed

when $n_2 \geq 1 > n_1$, x_2 is expanded

Compression and expansion

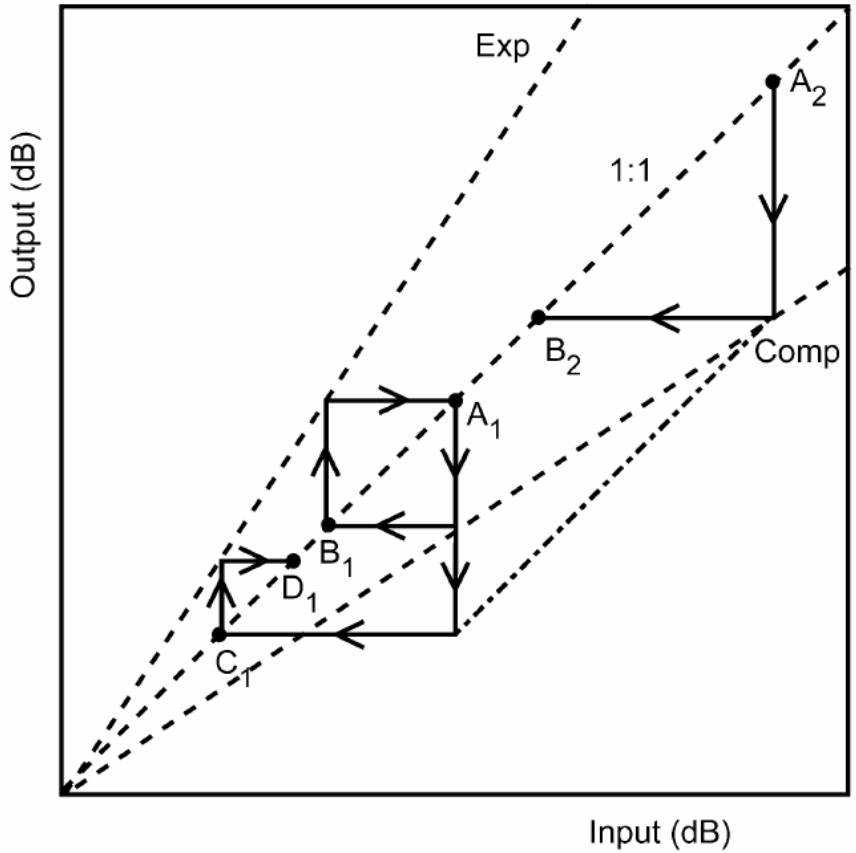


Fig. 3. Intuitive view of the companding strategy.

Output of companding

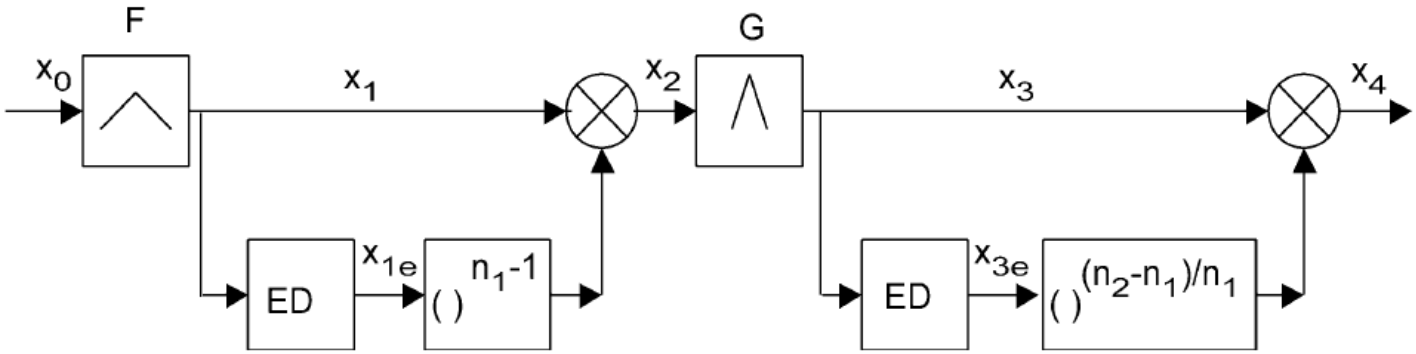


Fig. 2. Detailed view of a single channel of processing in Fig. 1.

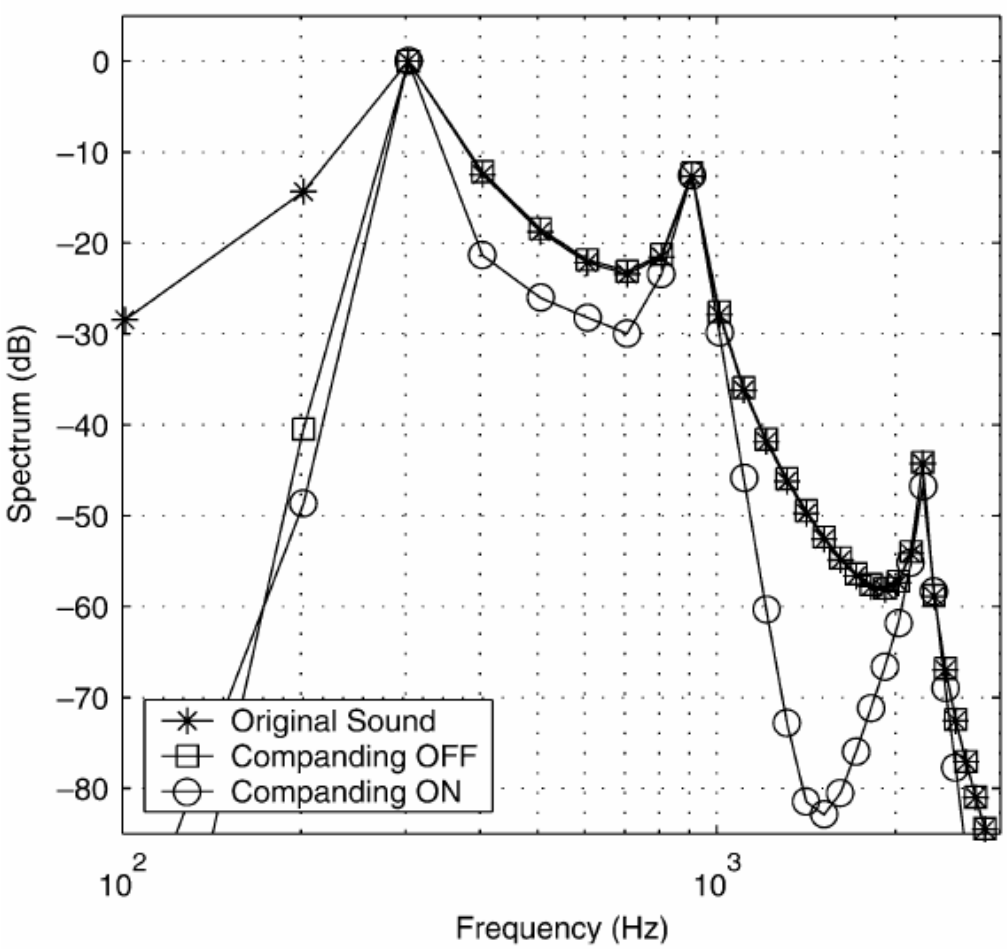
$$x_0 = a_1 \sin(w_1 t) + a_2 \sin(w_2 t + \varphi_0)$$

a_1, a_2 are the amplitude, w_1, w_2 are frequencies of two tones, φ_0 is phase

$$x_4 = \left[a_1 \left(\frac{a_1 + f_2 a_2}{a_1} \right)^{(n_1-1)/n_1} \right]^{n_2} \sin(w_1 t + \varphi_1 + \vartheta_1)$$

f_2 is the gain of filter F, φ_1 is the phase of filter F, ϑ_1 is the phase of filter G

Output of companding



ZCPA + companding

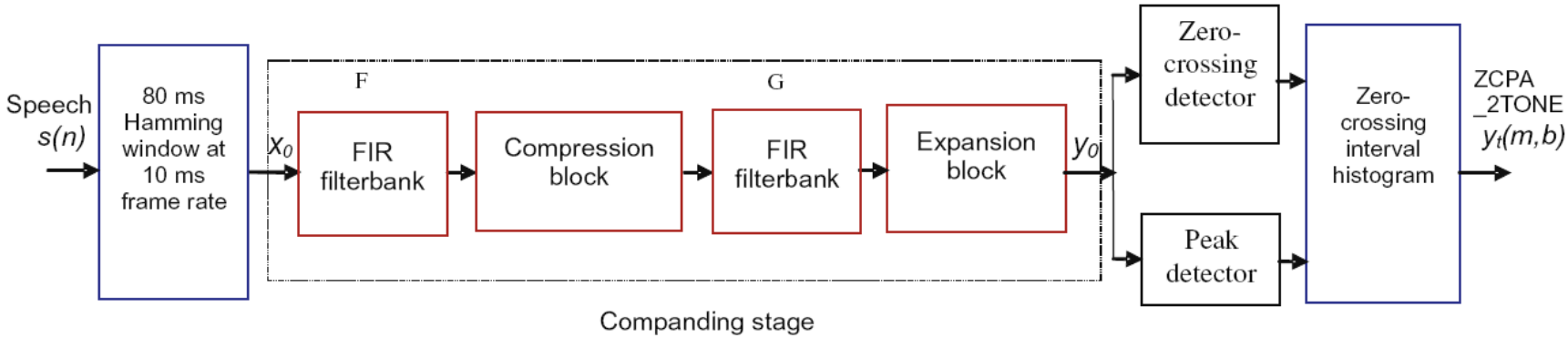
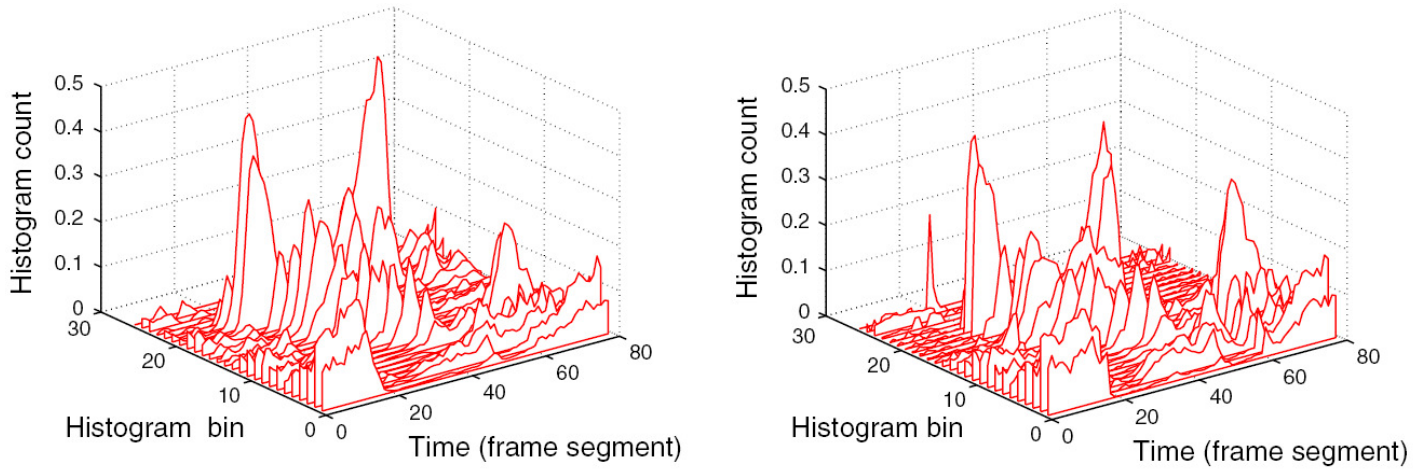


Fig. 10. The ZCPA with two-tone suppression as an ASR front end.

Time-frequency plot of ZCPA



(a) Base ZCPA (b) ZCPA with 2-tone suppression

Fig. 13. The time–frequency plots of the ZCPA histogram for (a) ZCPA and (b) ZCPA with two-tone suppressed output for the digit utterance 'one' in clean condition. The magnitude inhibition may be observed along the z-axis (vertical) at lower frequencies (higher histogram bins). (a) Base ZCPA; (b) ZCPA with 2-tone suppression.

experiments

- Corpus : speaker independent isolated digits from TIDIGITS.
- Four kinds of noise from NOISEX 92: white noise, factory noise, babble noise, Volvo noise.

Experiment results

Table 6
Continuous density HMM recognition rates (%) of the ZCPA with two-tone suppression compared with the base ZCPA and ZCPA with synaptic adaptation for isolated digits (TIDIGITS) with male speakers

SNR (dB)	White			Factory		
	ZCPA	ZCPA _ADP	ZCPA _2TONE	ZCPA	ZCPA _ADP	ZCPA _2TONE
Clean	95.4	95.4	95.4			
40	90.9	95.4	95.4	95.4	90.9	95.4
30	81.8	90.9	54.5	86.3	81.8	95.4
15	77.3	72.7	50.0	81.8	77.3	90.9
10	68.2	59.1	45.4	68.8	72.7	86.3
5	50.0	31.8	40.9	50.0	68.8	72.7
	Babble			Volvo		
40	90.9	90.9	95.4	95.4	95.4	95.4
30	86.3	86.3	90.9	90.9	90.9	95.4
15	72.7	72.7	81.8	86.3	86.3	95.4
10	63.6	59.1	68.1	86.3	81.8	95.4
5	31.8	54.5	54.5	81.8	77.3	90.9

Conclusions

- synaptic adaptation has a greater effect on high-frequency articulation and performs better in white Gaussian noise.
- two-tone suppression has a greater effect on low-frequency articulation and performs better in non-Gaussian real-world noise.