



SPEECH ENHANCEMENT FOR NOISE-ROBUST SPEECH RECOGNITION

Vikramjit Mitra and Carol Y. Espy-Wilson
Speech Communication Lab



Introduction

Propose an adaptive speech enhancement technique.

- Detects SNR of input speech.
- Removes stationary background noise.
- Performs pre-emphasis.
- De-noise using Modified Phase Opponency (MPO) model with Aperiodic Periodic and Pitch detector (APP)

Applications

- Noise-robust Speech Recognition
- Speech Enhancement
- Noise robust automated transcription

Corpus

- 'Speech in Speech shaped noise' (SSN) corpus of Speech Separation Challenge 2006 [3].
- Noise at 5 diff SNR: Clean, 6dB, 0dB, -6dB & -12dB
- Task: (a) detect key words, (b) detect all words

MPO-APP

MPO acts as a switch,

- When speech is dominant, it passes the signal as-is.
- When noise is dominant, it attenuates the signal.

❖ Problematic situations:

- (a) wide-band speech → speech deletions
- (b) narrow band noise → noise insertions
(i.e 2 formants close together)

❖ Solution: Use APP in cascade [2]

- APP detects periodic regions
- ✓ wide-band speech □ Periodic regions
- ✓ narrow-band noise □ Aperiodic regions

❖ APP thresholds vary with SNR

- Estimate SNR before MPO-APP enhancement

❖ MPO-APP fails to enhance high frequency formants if they are relatively weak

- Perform pre-emphasis before MPO-APP

Recognition Accuracy for MPO and MPO-APP

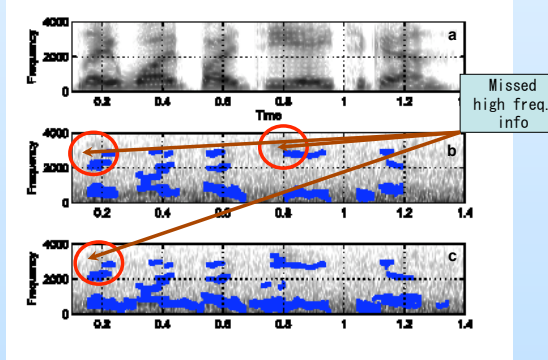
		∞ dB	6dB	0dB	-6dB	-12dB
no proc	CLD	98.56	56.67	18.94	11.78	11.67
	WRD	98.56	56.67	18.94	11.78	11.67
MPO [3]	CLD	96.00	73.83	50.06	26.00	14.33
	WRD	98.65	87.53	75.00	62.65	56.33
MPO-APP	CLD	95.94	75.06	53.17	31.22	19.06
	WRD	98.64	87.50	75.76	64.35	57.80

CLD: Color-letter-digit keyword recognition as specified in SSC 2006

WRD: Word recognition rate of the corpus

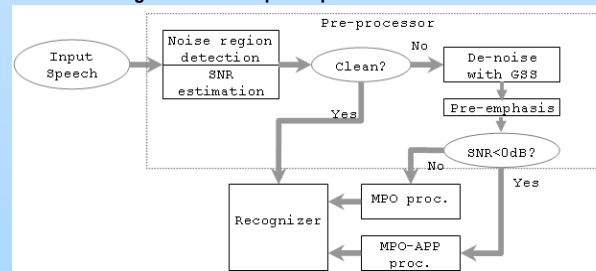
- MPO was implemented for SSC06 by Deshmukh et. al [1].

(a) Spectrogram of clean speech "lay white by d 8 again", (b) the MPO profile and (c) the MPO-APP profile of the same signal at -6dB SNR.

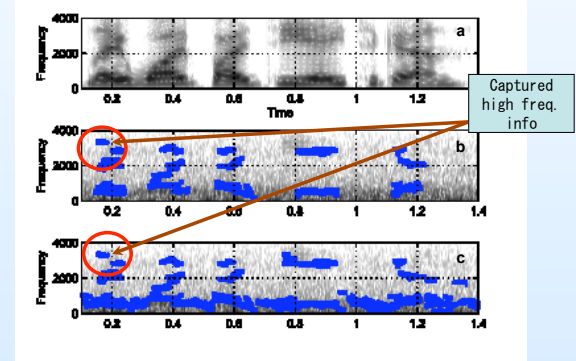


Solution

Block diagram of the adaptive speech enhancement scheme

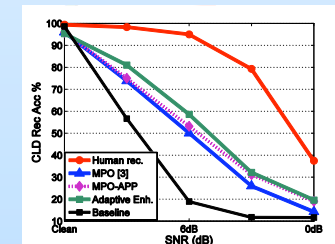


(a) Spectrogram of clean speech same as Fig.1, (b) pre-processed MPO & (c) pre-processed MPO-APP profile of the same signal at -6dB SNR.



Recognition accuracy for the proposed adaptive enhancement

		∞ dB	6dB	0dB	-6dB	-12dB
Adaptive enhancmnt	CLD	95.33	81.00	58.61	32.17	19.61
	WRD	98.35	91.06	79.11	66.02	58.53



Recognition accuracy for speech corrupted by speech shaped noise

Conclusion

- Proposed scheme increases CLD accuracy by 17% (mean) and WRD accuracy by 3.7% over MPO [1].
- Average increase by 98.04% over baseline.
- Reduces unwanted-computation.

References

- [1] O.D Deshmukh and C. Espy-Wilson, "Modified Phase Opponency Based Solution to the Speech Separation Challenge", In Proc. of Interspeech 2006, pp. 101-104, Pittsburgh, PA.
- [2] O.D. Deshmukh, C. Espy-Wilson and L.H. Carney, "Speech Enhancement Using The Modified Phase Opponency Model", Journal of Acoustic Society of America, Vol. 121, No. 6, pp 3886-3898, 2007.
- [3] M. Cooke, J. Barker, S. Cunningham and X. Shao, "An audio-visual corpus for speech perception and automatic speech recognition", Journal of Acoustic Society of America, Vol. 120, pp 2421-2424, 2006.