A Soft Masking Strategy for Simultaneous Suppression of Noise and Reverberation in Cochlear Implants

Oldooz Hazrati, Seyed Omid Sadjadi, Hussnain Ali, Philipos C. Loizou, and John H. L. Hansen

Department of Electrical Engineering, The University of Texas at Dallas, Richardson, TX, USA

{hazrati, sadjadi, hussnain.ali, john.hansen} @ utdallas.edu

1. Introduction

- Speech recognition performance by cochlear implant (CI) users degrades exponentially in reverberant environments.
- Unlike reverberation, noise is additive and affects speech in a different and complimentary fashion.
- Noise masks the weak consonants to a greater degree than the higher intensity vowels, but unlike reverberation this masking does not depend on the energy of the preceding segments.
- Hence, the combined effects of reverberation and noise adversely affect speech intelligibility more than either reverberation or noise alone.
- A single-channel non-ideal solution to the problem of noisy reverberant speech enhancement for CI users is proposed in the present study.

2. Noise and Reverberation Suppression

- The noise power spectral density (PSD) is computed from the first 100 ms of the corrupted signal (no convolutive distortions due to reverberation exist).
- The PSD of late reflections can be modeled as a delayed and smoothed version of the PSD of reverberant speech as:

$$|R_l(t,f)|^2 = \alpha\, w(t-t') * |X(t,f)|^2$$

- The superposition of noise and late reverberation PSDs is considered as the PSD of distortion (caused by both reverberation and noise):

$$|D(t,f)|^2 = |R_l(t,f)|^2 + |\hat{Y}(t,f)|^2$$

- The soft mask for time frame ***t*** and frequency bin ***f*** is computed as:

$$M(t,f) = \left(\frac{\lambda(t,f)}{\lambda(t,f)+\delta}\right)^{\theta}$$

A priori signal-to-distortion ratio (SDR)

$$\lambda(t,f) = \mu\cdot\lambda(t-1,f) + (1-\mu)\cdot\frac{|X(t,f)|^2}{|D(t,f)|^2}$$

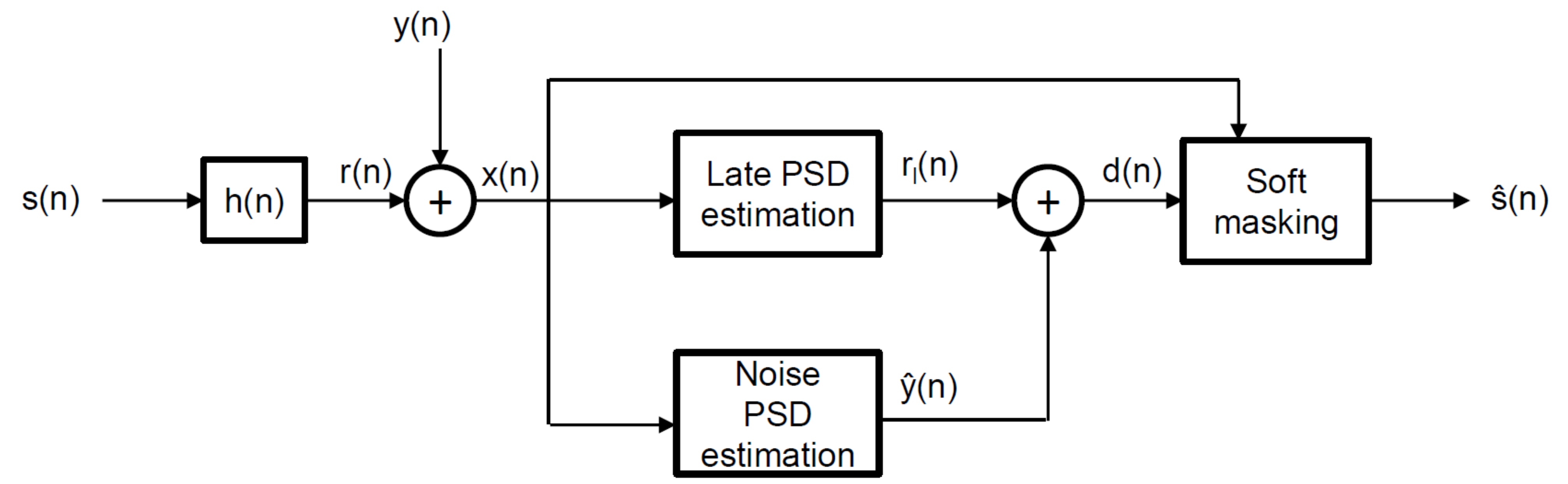

Fig. 1. Block diagram of the proposed masking strategy

3. Method

- A total of 7 CI users participated in the listening tests.
- To generate the reverberant stimuli, IEEE sentences were convolved with impulse responses of a 10.06 m x 6.65 m x 3.4 m room with reverberation times (T_{60}) of 0.6 s and 0.8 s. Speech-shaped noise was added to the reverberant signals at RSNR = 15 dB.
- Two IEEE lists (20 sentences) were used for each condition and the condition order was randomized across subjects.

4. Results

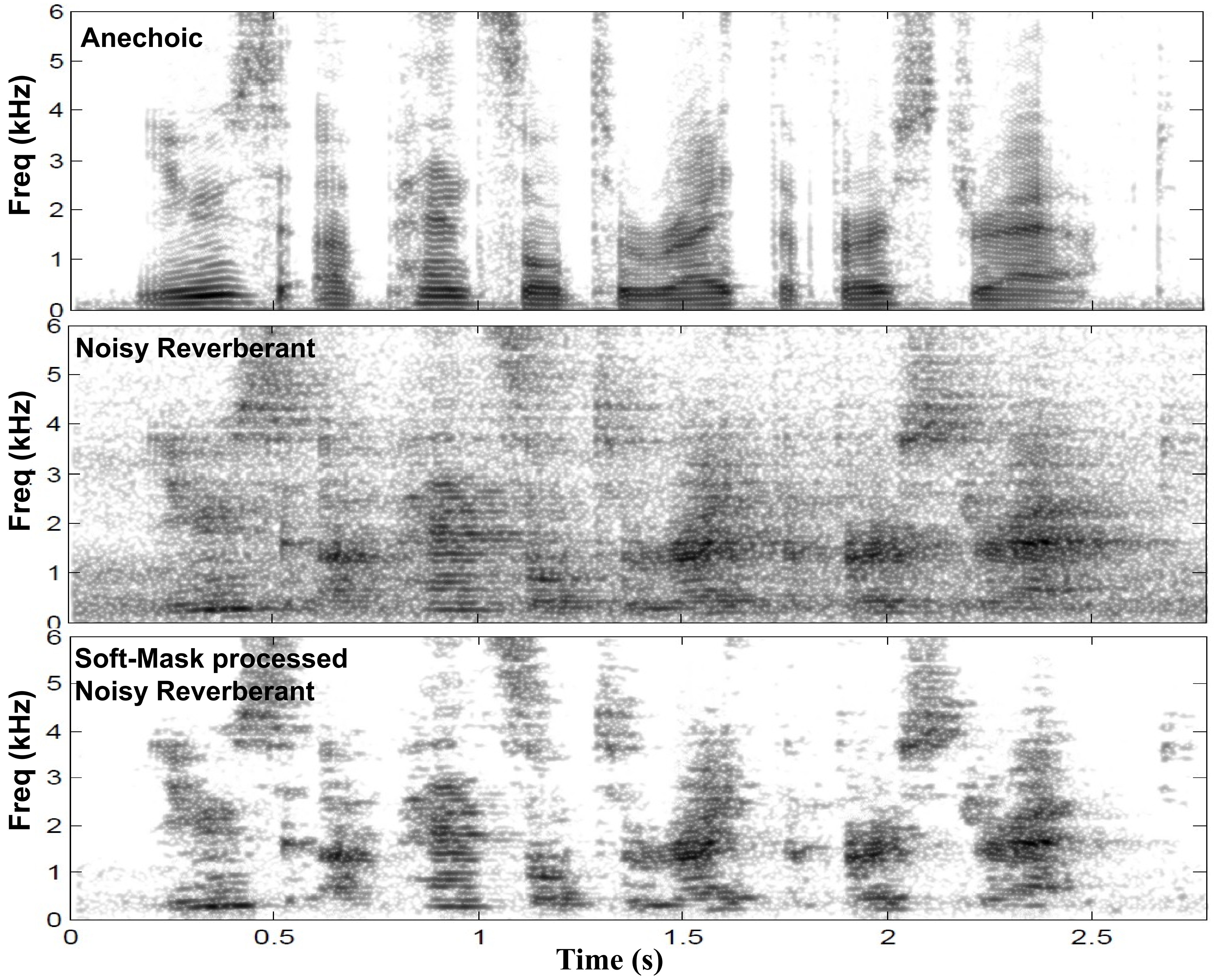

Fig. 2. Spectrograms of IEEE sentence "use a pencil to write the first draft" T_{60} = 0.6 s and RSNR = 15 dB

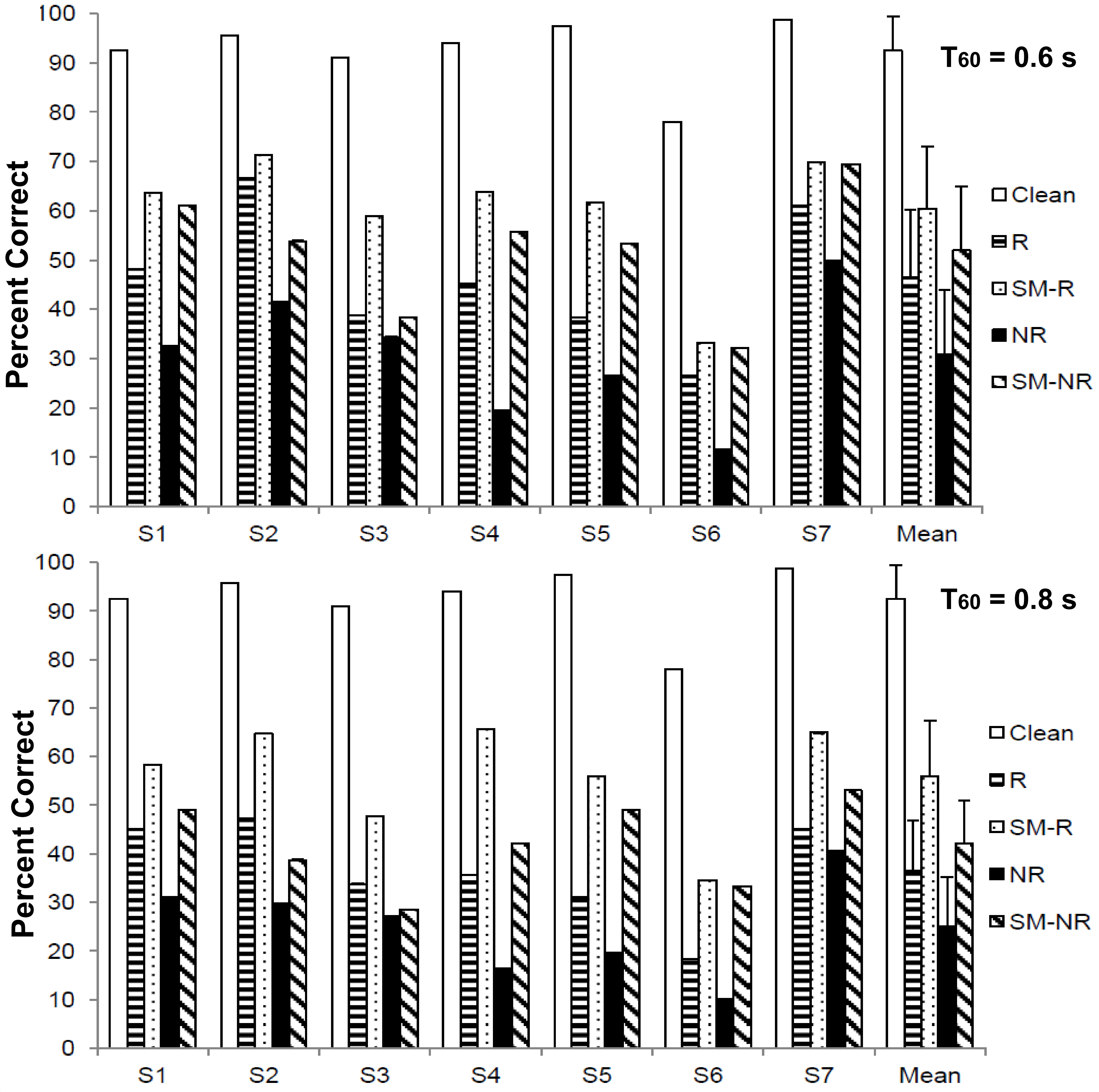

Fig. 3. Individual intelligibility scores

- The speech intelligibility scores improved from an average of 46.42% and 36.59%, to 60.35% and 55.98% at T_{60} = 0.6 and T_{60} = 0.8 s conditions, respectively.
- In the presence of noise, speech identification scores improved from an average of 30.89% and 24.96% to 51.97% and 41.98%, respectively.

5. Conclusions

- ANOVA (with repeated measures) confirmed a significant effect ($F[1,6]$ = 25.59, p < .005) of T_{60}, a significant effect of noise ($F[1,6]$ = 31.19, p < 0.005) and a significant effect of processing ($F[1,6]$ = 47.03, p < .005) on speech intelligibility.
- Results indicated that in both reverberant-alone and noisy reverberant conditions (T_{60} = 0.6 and 0.8 s), intelligibility of the processed stimuli improved significantly (p < 0.0001, paired samples t-tests, Bonferroni corrected) compared to the unprocessed reverberant and noisy reverberant stimuli.

Research supported by NIDCD/NIH and Cochlear Ltd.