Effect of distortion on the intelligibility of speech in noise

Julián Villegas

University of Aizu, 2024

Distortion

- **Linear distortion:** Alteration of magnitude (and phase) but no creation of new spectral content (Example, dynamic-range compression)
- **Non-linear distortion:** Creation of spectral content (clipping, spectral shaping, etc.)

Problem

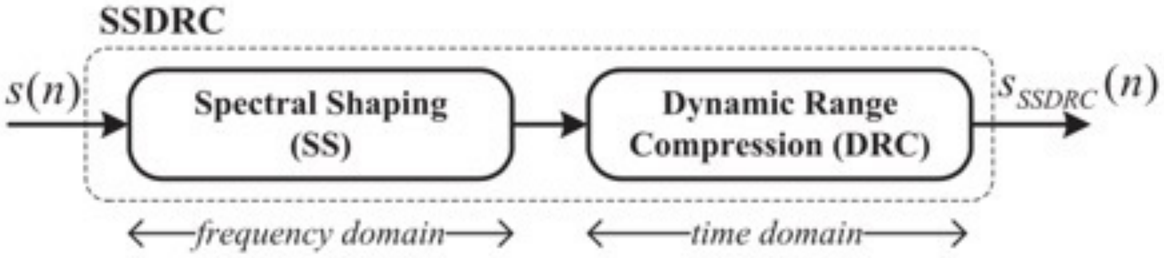
- When noise is present, it’s difficult to hear what’s being said over PA systems, for example.
- This is usually solved by increasing the volume of the message:

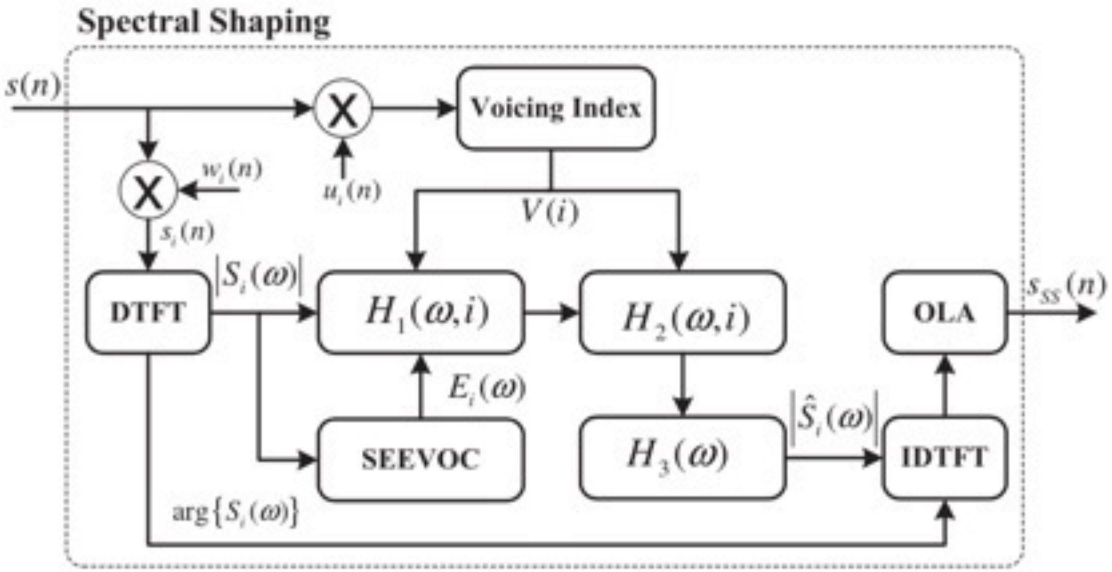
Synthetic speech

Live speech

- But, increasing the sound level is dangerous for health [1].
- So, DSP solutions have been proposed to increase speech intelligibility without increasing volume

Spectral shaping & dynamic range compression [2]





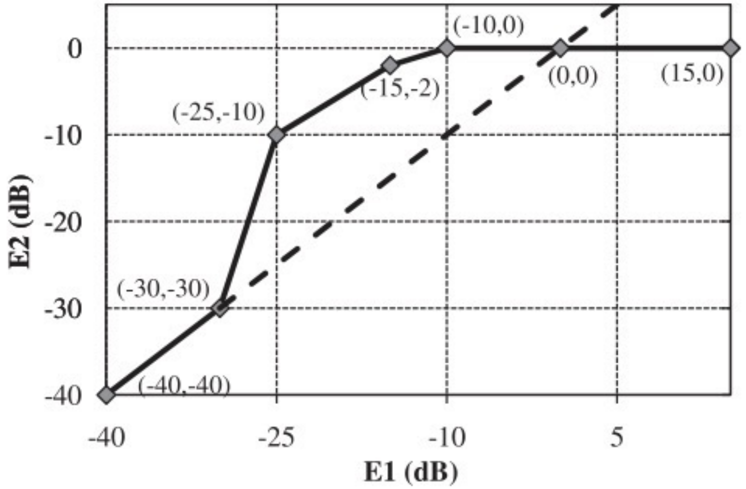
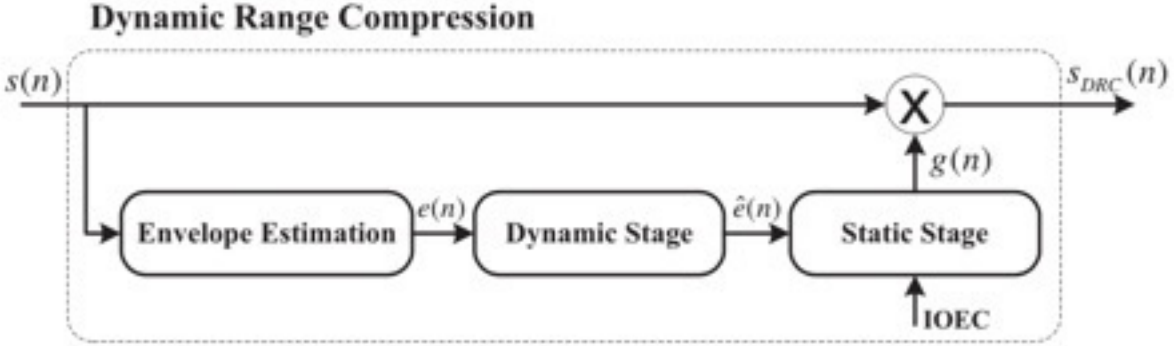
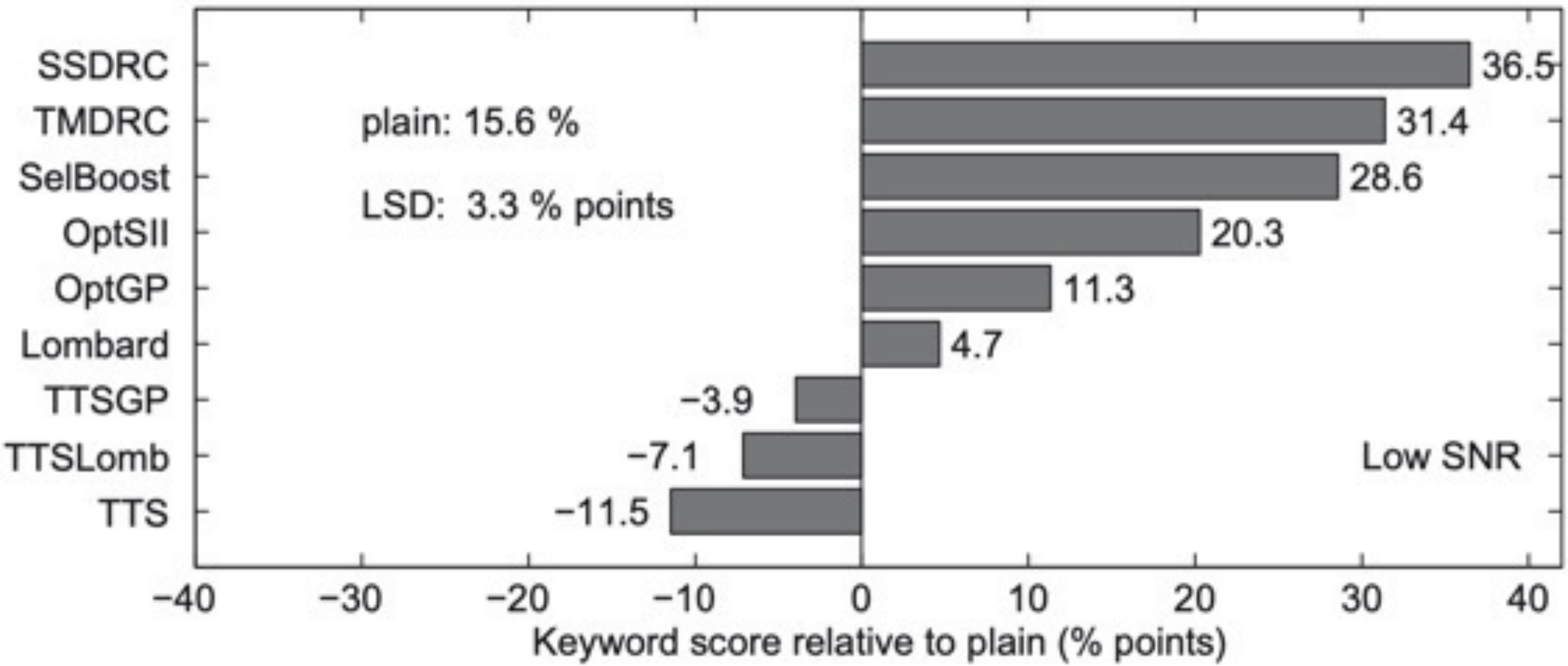


Fig. 5. IOEC of the static stage of DRC.

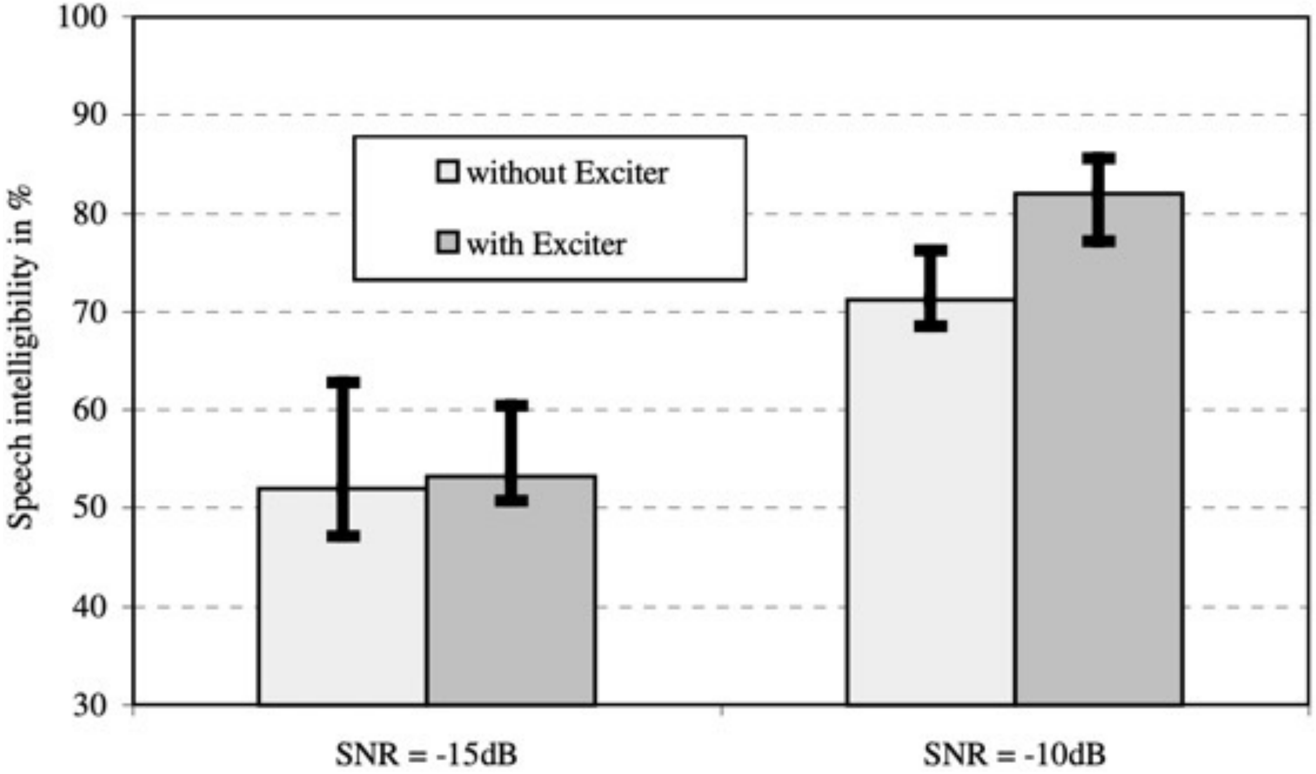


Subjective intelligibility of SSDRC [3]



Aural exciter and speech intelligibility [4]

- Note that aural exciter is NOT a state or the art on speech intelligibility in noise
- It is a general-purpose audio effect
- Intelligibility was investigated by Chalupper in [4] and because of that is mentioned here
- His results were inconclusive



Masker: Speech Modulated Noise (SSN)

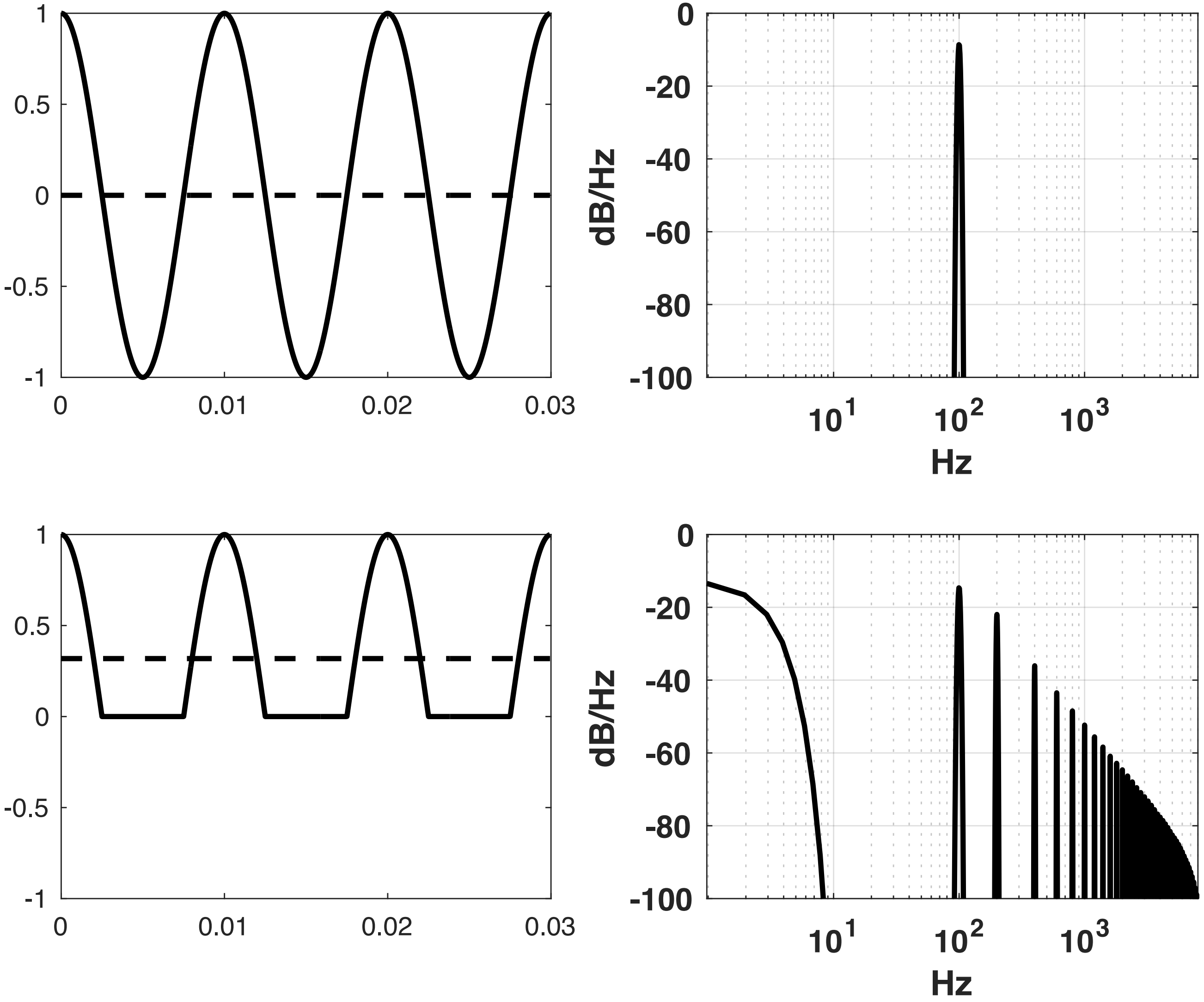
Half-wave rectification

$$y(t) = \frac{1}{\pi} + \frac{1}{2}\cos(\omega_0 t) - \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{(2n)^2 - 1}\cos(2n\omega_0 t). \quad (1)$$

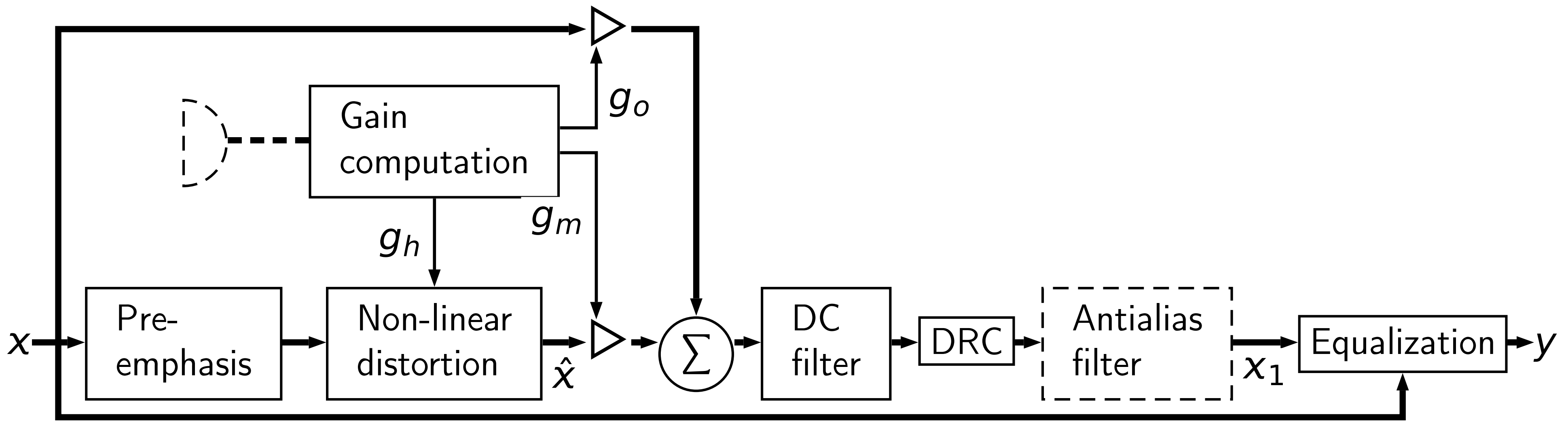
- Half-wave rectification: negative values -> 0
- Preserves original fundamental frequency
- Increase harmonic contents (non-linear distortion)
- Increase the DC offset
- Potentially create aliasing frequency components

Effect of halfway rectification

- The figure shows the effects for a 100 Hz cosine-wave
- Left panels are the time signals
- Right panels are their spectra
- The dashed line is the DC-offset
- Halfway rectification created a positive DC-offset
- For best operation of a DRC, the wave should be centered around zero, so a DC-filter is needed after halfwave rectification



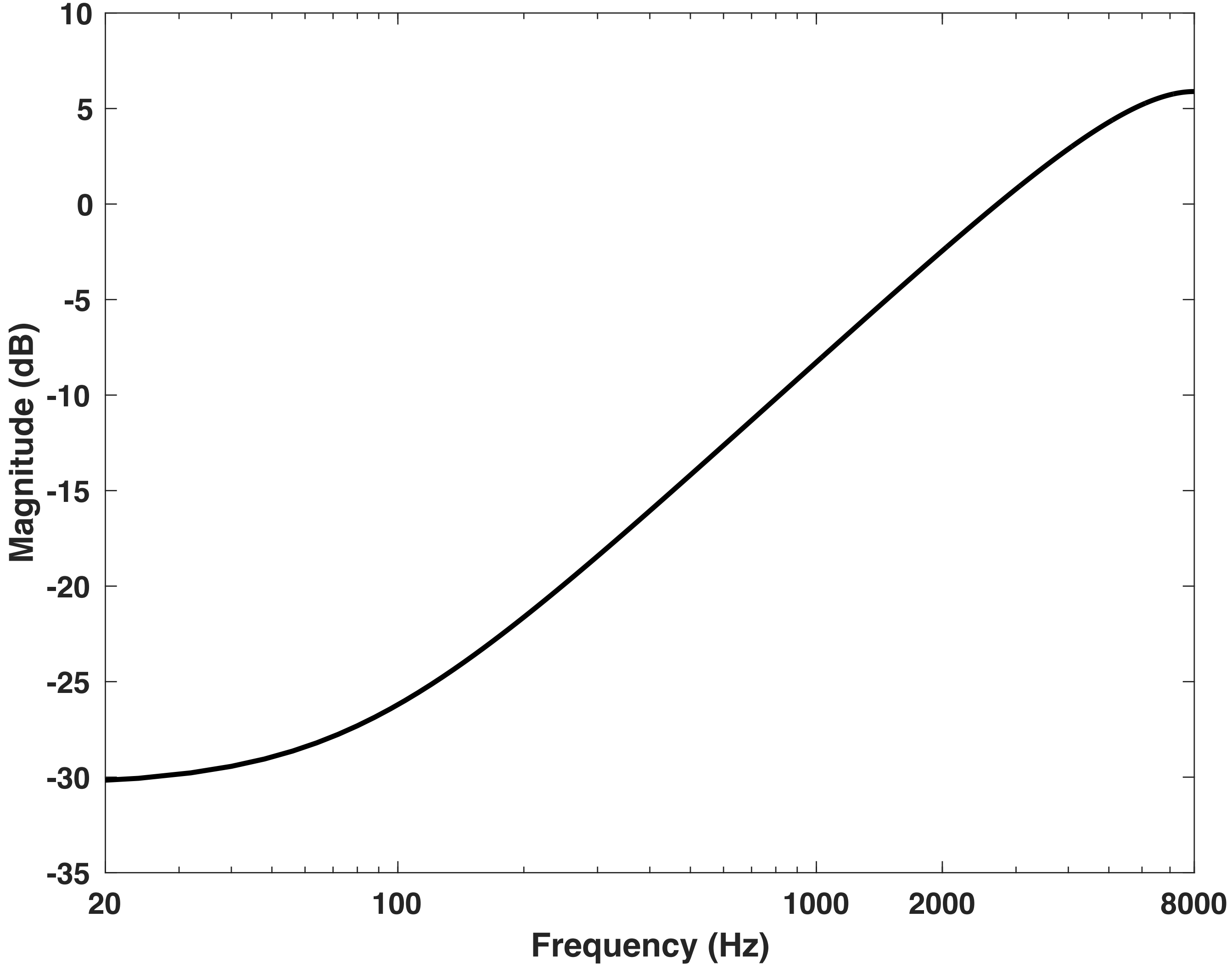
Basic proposal



- Pre-emphasis: emphasizes high-frequencies, attenuates low ones.
- DC-filter: Eliminates any DC component in the signal
- DRC: Dynamic Range Compression (linear distortion)
- Equalization: apply a gain to the new formed signal so that it has the same RMS as the original
- Antialiasing: suppress energy above Nyquist frequency.

Pre-emphasis filter

- The pre-emphasis filter is a standard procedure that attenuates low frequencies and exhacerbates higher ones, as shown in the figure

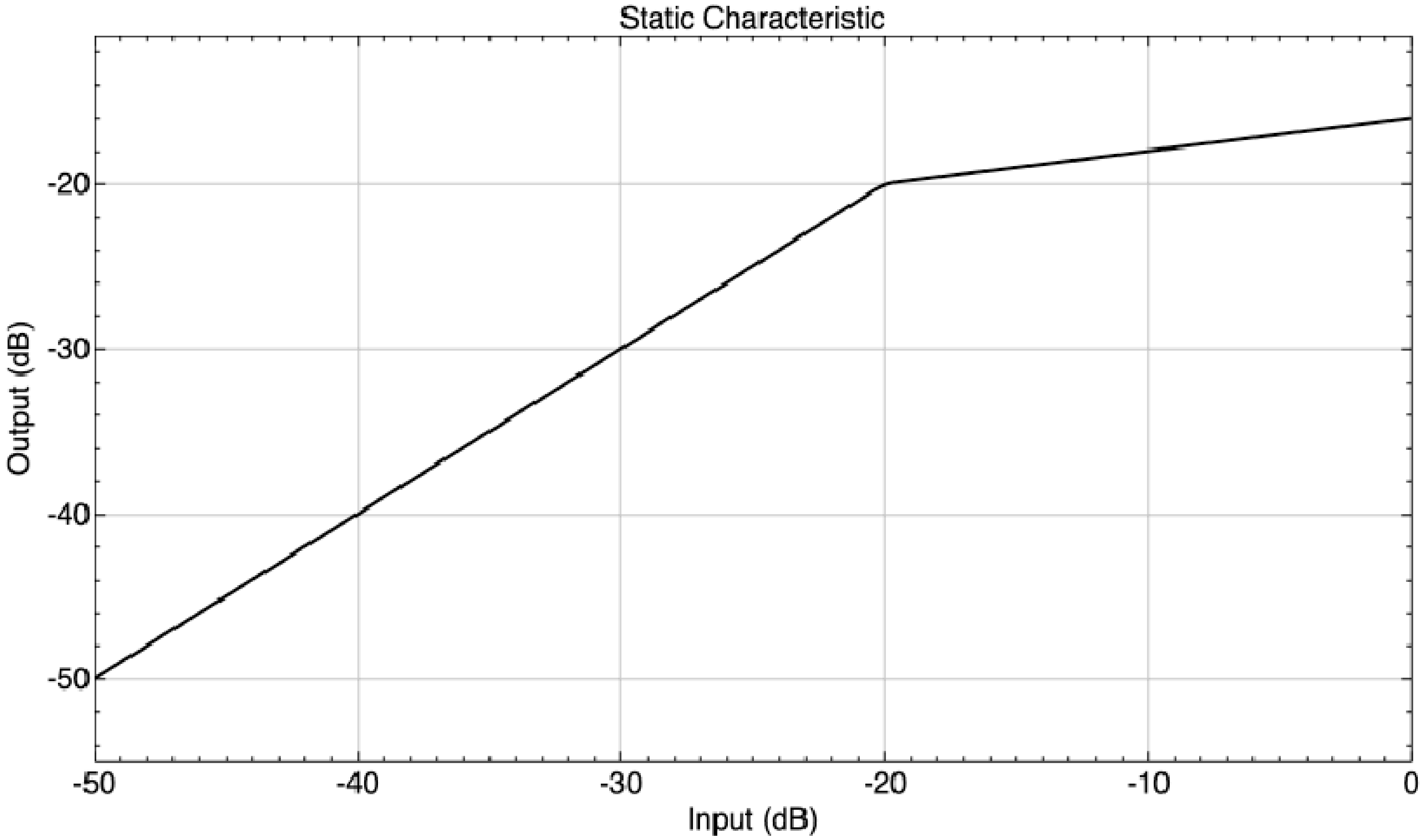


Dynamic Range Compression

- The range between the loudest and quietest part of an audio signal is known as its dynamic range.
- It is often desirable to boost quieter parts of an audio signal while keeping the louder parts at bay so that they do not exceed the amplitude constraints of the audio system (clipping). In those cases, DRC can be used.
- DRCs have different parameters, as a first approximation these values were used:
- **Threshold** (the point at which the compression starts): -20 dB
- **Ratio**: 5:1 (5 dB in the input will be represented by 1 dB in the output)
- **Attack** (the time to start working after threshold has been reached): 1 ms
- **Release** (the time to stop working after threshold has been reached): 5 ms
- **Output** (additional gain, if desired): +0 dB
- **Knee** (spread around the threshold to make it smooth) = 1 dB

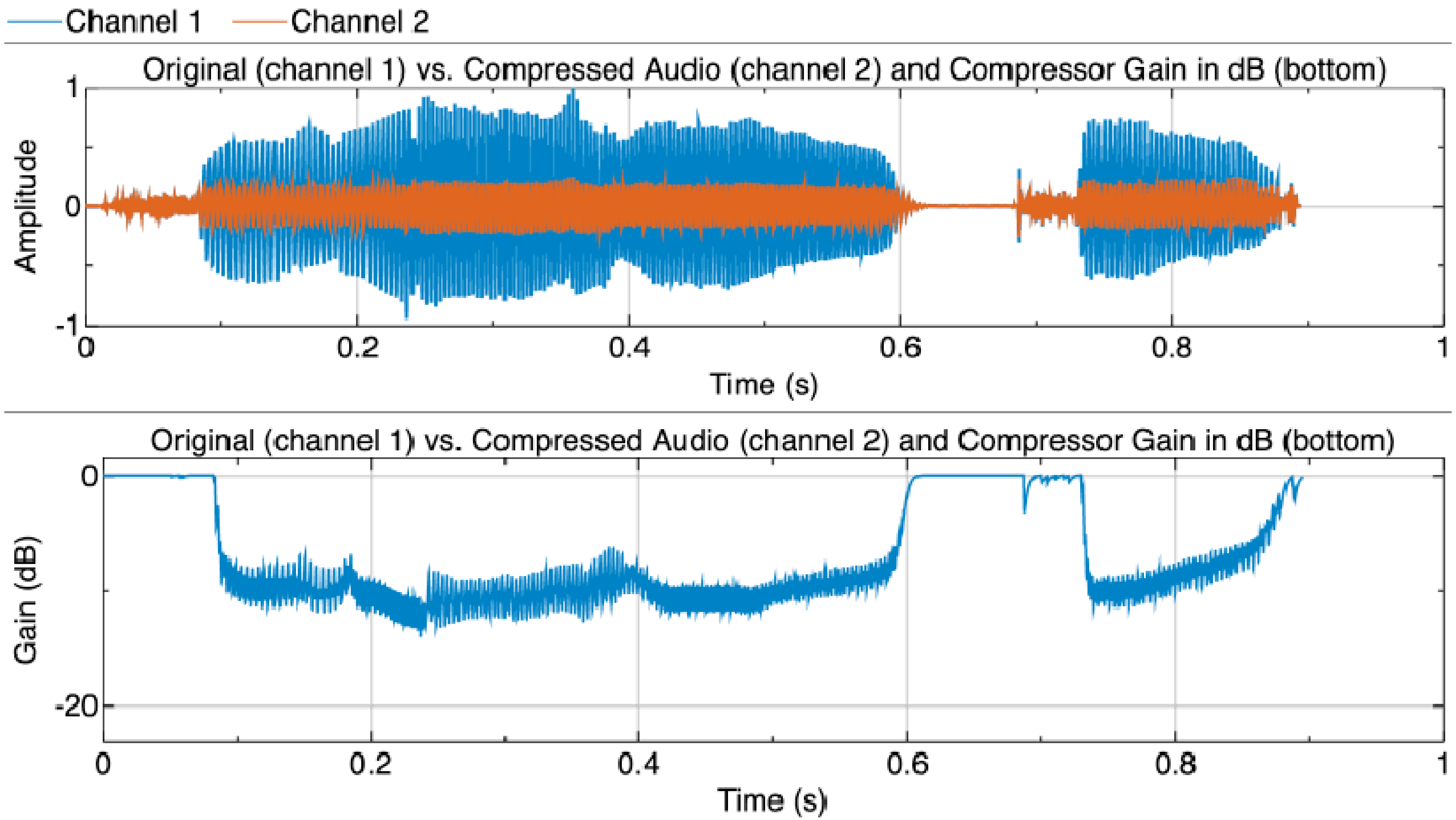
Static characteristic curve

- With the default values for the parameters shown in the previous slide, the static curve shown here is obtained
- This is simpler than in SSDRC (slide 4)

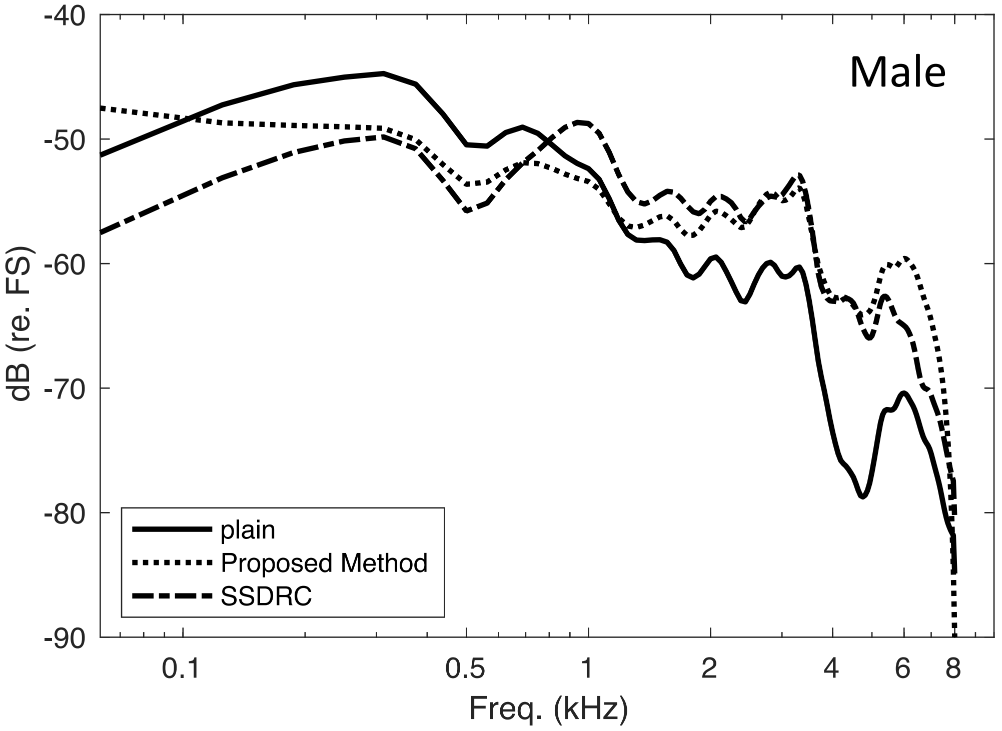


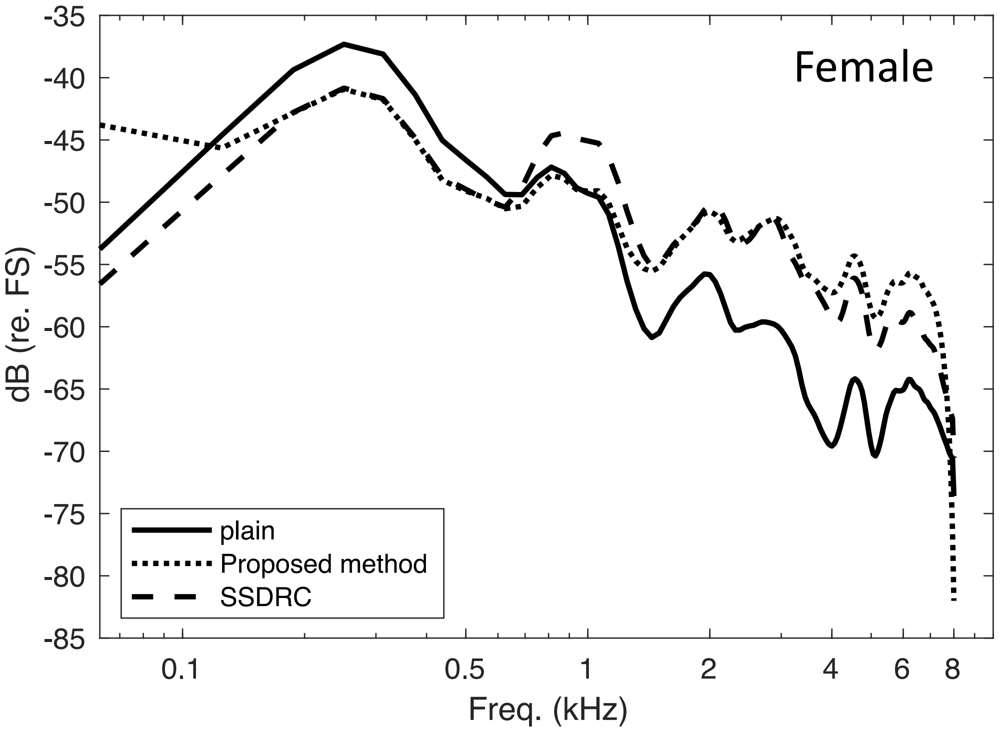
DRC effect on speech

- The resulting wave is in general quieter
- Once maximized the difference between quiet and loud speech is smaller than in the original



Effect of the proposed method on the LTAS spectrum



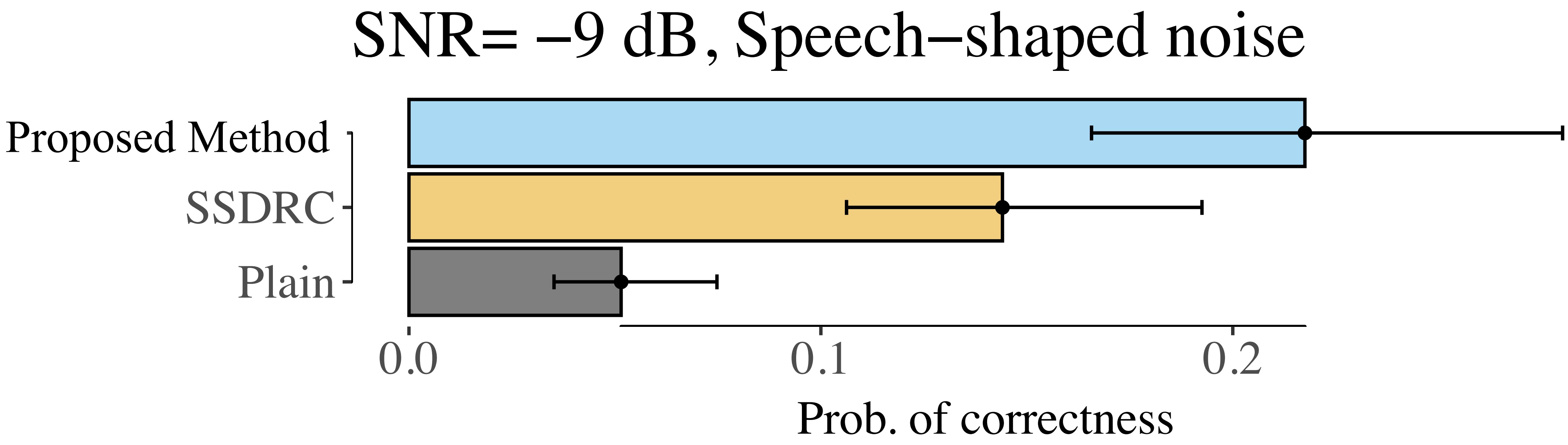


Timing results (CPU time)

Method	time	rel. time
SSDRC	431,426.45	100%
HDDRC	20,562.46	4.8%
O_HDDRC	19,911.96	4.6%

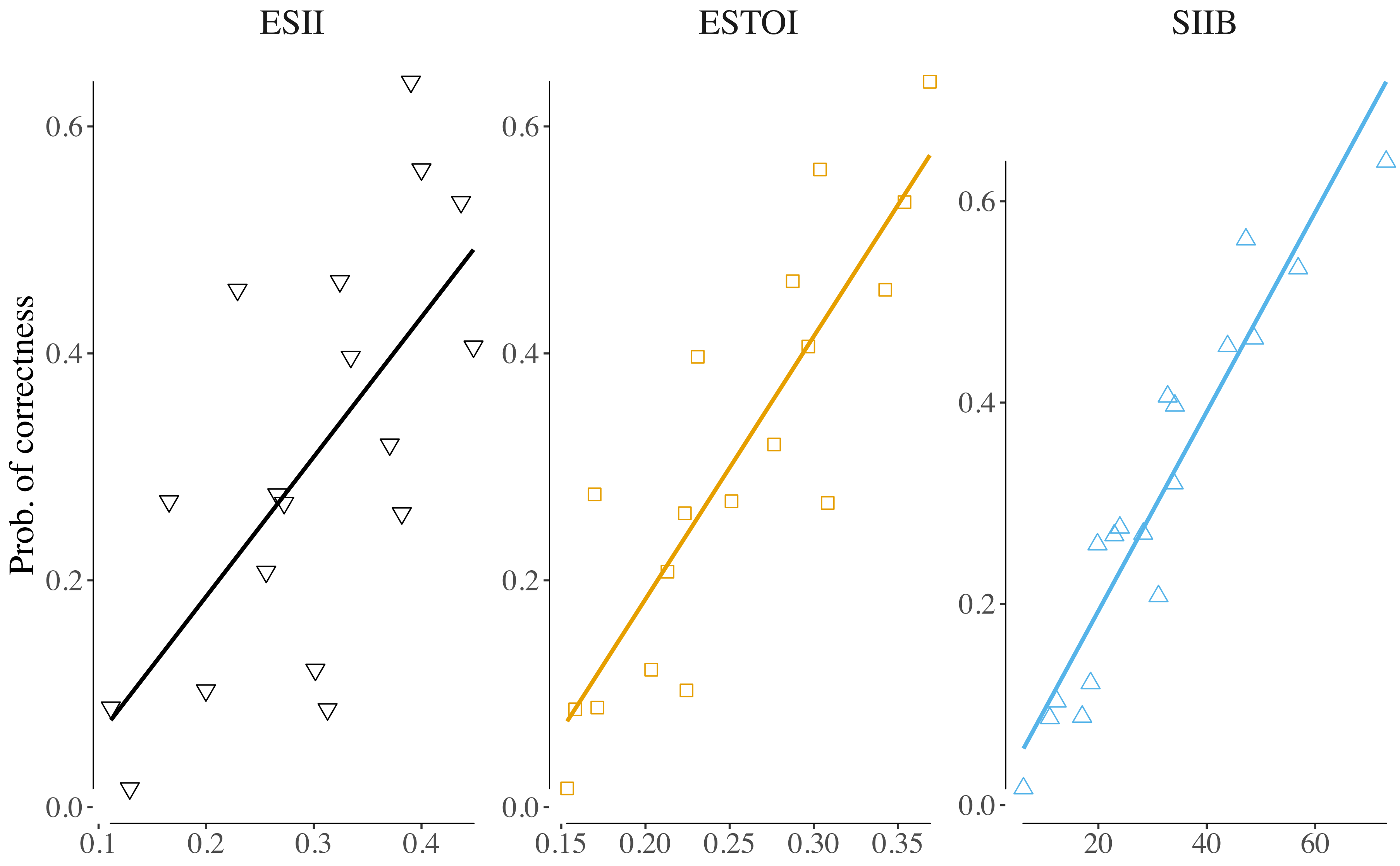
- The proposed system works only in the time domain
- Some components can be re-utilized to speed up the computation (O_HDDRC)

Final subjective results



Correlation with objective metrics [6]

ESII 0.683
ESTOI 0.869
SIIB 0.947



Conclusions

- The proposed method is similar to some existing methods
- However, the combination here proposed is unique
- It works by increasing the level in regions important for speech
- The proposed method is fast (it can work only in the time domain)
- The proposed method yields better intelligibility gains or similar gains than SS-DRC (as shown in the subjective experiment)
- There is room from improvement (e.g., the parameters of the DRC)
- Several variations are possible

- [1] WHO, “Burden of disease from environmental noise, Quantification of healthy life years lost in Europe.” World Health Organization, 2011.
- [2] T.-C. Zorila, Y. Stylianou, T. Ishihara, and M. Akamine, “Near and far field speech-in-noise intelligibility improvements based on a time–frequency energy reallocation approach,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 10, pp. 1808–1818, 2016.
- [3] M. Cooke, C. Mayo, C. Valentini-Botinhao, Y. Stylianou, B. Sauert, and Y. Tang, “Evaluating the intelligibility benefit of speech modifications in known noise conditions,” Speech Communication, vol. 55, no. 4, pp. 572–585, 2013.
- [4] J. Chalupper, “Aural exciter and loudness maximizer: What’s psychoacoustic about “psychoacoustic processors”?,” in Proc. 109 Audio Eng. Soc. Conv., 2000.
- [5] K. Wojcicki, K. Fitz, K. Recker, D. Reynolds, and T. Zhang, “Sidechain harmonic enhancement of noise corrupted speech for hearing impaired listeners,” in IEEE Wkshp. on Applications of Signal Processing to Audio and Acoustics (WASPAA), pp. 1–5, 2015. DOI: 10.1109/WASPAA.2015.7336926.
- [6] S. Van Kuyk, W. B. Kleijn, and R. C. Hendriks, “An evaluation of intrusive instrumental intelligibility metrics,” IEEE/ACM Trans. on Audio, Speech, and Language Process., vol. 26, no. 11, pp. 2153–2166, 2018.