ROBUST PARAMETER ESTIMATION FOR AUDIO DECLIPPING IN NOISE

Carnegie Mellon

Mark J. Harvilla and Richard M. Stern
Department of Electrical and Computer Engineering
Carnegie Mellon University, Pittsburgh, PA, USA

Electrical & Computer ENGINEERING

Abstract

Contemporary audio declipping algorithms often ignore the possibility of the presence of additive channel noise. If and when noise is present, however, the efficacy of any declipping algorithm is critically dependent on the accuracy with which clipped portions of the signal can be detected. This paper introduces an effective technique for inferring the amplitude and percentile values of the clipping threshold, and develops a statistically optimal classification algorithm for accurately differentiating between clipped and unclipped samples in a noisy speech signal. The overall effectiveness of the clipped sample estimation algorithm is evaluated by the degree to which automatic speech recognition performance is improved when decoding speech that has been declipped with state-of-the-art declipping algorithms paired with the clipped sample estimation algorithm. Up to 35% relative improvements in word error rate have been observed. Beyond the accuracy of the developed techniques, this paper generally underscores the necessity of robust parameter estimation methods for declipping in noise.

Audio Clipping and Noise

$$x_c[n] = \begin{cases} x[n] & \text{if } |x[n]| < \tau \\ \tau \cdot \operatorname{sgn} x[n] & \text{if } |x[n]| \geq \tau \end{cases}$$

Audio clipping typically occurs in one of three ways:

1. Upon recording, as a result of exceeding the dynamic range limitations of the A/D converter.
2. As a result of writing improperly-amplitude-normalized data to a file.
3. On purpose, to achieve a desirable perceptual characteristic.

This paper concerns the estimation of which samples were clipped when the clipped signal, $x_c[n]$, itself is corrupted by additive Gaussian noise, $w[n]$:

$$y[n] = x_c[n] + w[n]$$

This situation may occur when a speech signal is clipped on capture, then further degraded by environmental and channel noise during transmission.

Clipping Threshold Estimation

- The presence of clipping in a noisy speech signal can be determined through analysis of the signal's amplitude distribution. It manifests as two distinct peaks symmetric about 0.
- Using a simple peak finding algorithm, the locations of the K peaks of the smoothed amplitude distribution can be found; denote these K peaks : $\{k_0, k_1, k_2, \ldots, k_{K-1}\}$.
- An estimate of the clipping threshold is given by:

$$\tilde{\tau} = \frac{1}{K-1} \sum_{i=0}^{K-1} |k_i|$$

- Ideally, if no clipping is present, K=1, as there is only a single peak near 0, and the threshold estimate becomes ∞. If clipping is present, K=3, but the sum is divided by two, as one of the peaks remains very near 0 and thus does not contribute to the sum.

Threshold Percentile Estimation

- An estimate of the probability of a given noisy sample being clipped is related to the probability of an input signal sample being clipped. This can be determined from the percentile value of the clipping threshold, calculable as follows:

$$\text{Percentile value of } \tau = \int_{-\tau}^{+\tau} c(x)dx = \int_{-\infty}^{+\tau} c(x)dx - \int_{-\infty}^{-\tau} c(x)dx = C(\tau) - C(-\tau)$$

- In the above, $c(x)$ is the signal amplitude PDF, $C(x)$ is the CDF.

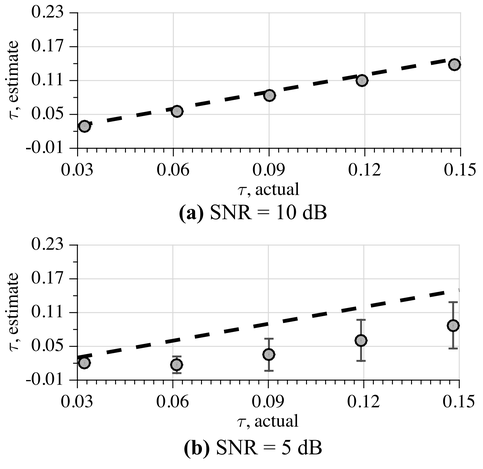
Clipped Sample Estimation

- The classification of observed noisy speech samples as either clipped or unclipped is done on a sample-by-sample basis, with each sample treated independently. A sample is classified as clipped if the following condition holds:

$$\frac{\Pr\left(x_c[n] = \pm\tau | y[n], \sigma_y^2, \sigma_w^2, \tau\right)}{\Pr\left(x_c[n] \neq \pm\tau | y[n], \sigma_y^2, \sigma_w^2, \tau\right)} \geq 1$$

- As described in more detail in the paper, the numerator and denominator of this ratio can be computed as functions of the percentile value of τ and the posterior probability of the observed value, $y[n]$, given that $x_c[n]$ is (not) equal to $\pm\tau$.
- The posterior probability of $y[n]$ given that $x_c[n]$ is not equal to $\pm\tau$ is modeled as a Gaussian distribution with zero mean and variance equal to the overall signal power.
- The posterior probability of $y[n]$ given that $x_c[n]$ is equal to $\pm\tau$ is modeled as a Gaussian distribution with mean $\pm\tau$ and variance equal to the noise power.

Experimental Results



(a) SNR = 10 dB

(b) SNR = 5 dB

Figure 1: Results of blindly predicting τ using peak estimation with Gaussian noise added after clipping.

Experimental Results (cont'd)

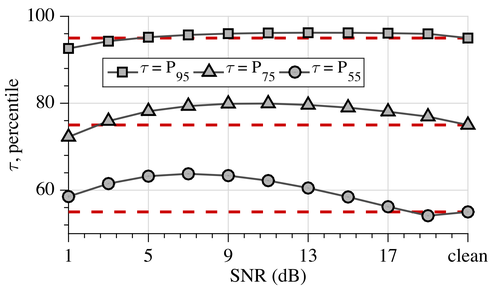


Figure 2: Results for predicting the value of τ using the signal amplitude CDF, given the amplitude value of τ.

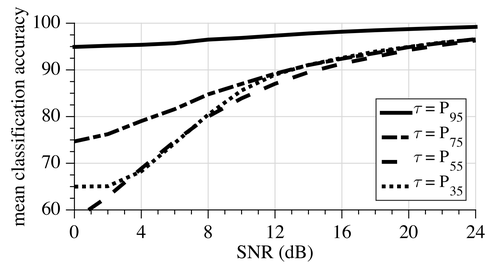


Figure 3: Mean clipped sample estimation accuracy in varying levels of noise and varying clipping thresholds.

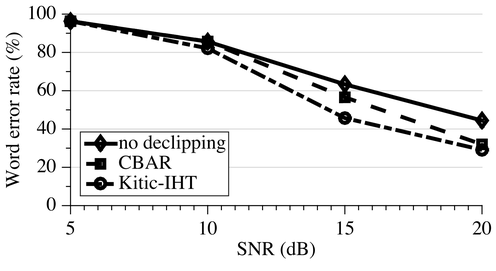


Figure 4: WER with two state-of-the-art declipping algorithms using wholly inferred information about which samples were clipped. Here, 25% of the signal samples were clipped before noise was added at the indicated SNR.

Summary

- Blind inference of clipping is necessary for practical declipping applications in noise.
- A systematic approach to clipping detection was presented and demonstrated to work in conjunction with state-of-the-art declipping techniques.