International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:8, No:2, 2014

Detection of Clipped Fragments in Speech Signals

Sergei Aleinik, Yuri Matveev

Abstract**—**In this paper a novel method for the detection of clipping in speech signals is described. It is shown that the new method has better performance than known clipping detection methods, is easy to implement, and is robust to changes in signal amplitude, size of data, etc. Statistical simulation results are presented.

Keywords**—**Clipping, clipped signal, speech signal processing.

I. Introduction

CLIPPING is a kind of signal distortion. The amplitude of a clipped signal is limited by some threshold(s). On oscillograms, clipping usually appears as a cutoff of signal amplitude. Clipping can be single-sided (only the top or only the bottom of the signal is cut) and double-sided. In digital clipped signals, the signal samples are grouped around their maximum and minimum values ("soft" clipping), or are simply equal to their corresponding maximum and minimum values ("hard" clipping). Mathematically the process of double-sided hard clipping of a discrete signal $x(i)$ can be written as follows [1]:

$$x_{cl}(i) = \begin{cases} x(i), & \text{if } |x(i)| < A \\ A(x(i)/|x(i)|), & \text{if } |x(i)| \geq A \end{cases}, \quad (1)$$

where i is a discrete time index, $x_{cl}(i)$ is the clipped signal and A is the clipping threshold.

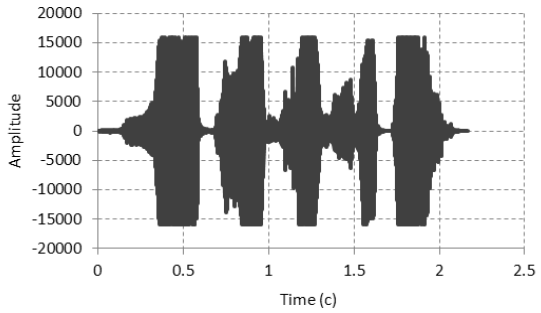

Fig. 1 Fragment of a double-sided clipped speech signal.

A typical view of a double-sided clipped speech signal for $A = 16000$ is depicted in Fig. 1. Clipping leads to an increase in high frequency signal components and to the appearance of higher order harmonics, which causes the deterioration of sound quality. Ultimately, clipping leads to the performance of speech processing systems (for example, speech recognition or speaker verification systems) to drop notably. Thus, the task of detecting clipped speech signal fragments (for the purpose of further rejection, for example) is quite important.

If we know the clipping threshold A and the signal power P_{sig}, we can calculate the so-called "clipping ratio" (CR), i.e. the characteristic of "how badly clipped the signal is" [1], [2]:

$$CR_{lin} = A/\sqrt{P_{sig}}$$
$$CR_{dB} = 10\log_{10}(A^2/P_{sig}) \quad (2)$$

In real life, however, the clipping threshold is unknown. So in (2) it is necessary to use the corresponding estimated values of A and P_{sig}, which leads to high variance of calculated CR.

Some articles [1], [2] on clipped signal detection are devoted to a narrow range of specific signals (e.g. OFDM signals) and use the specific characteristics of these signals.

Some algorithms use the knowledge of source (not clipped) signals [3] and therefore focus more on evaluating the quality of processing devices (e.g. amplifier, etc.) rather than signals themselves.

The purpose of this paper is to present additional research based on our new method, first described in [4]: a method of measuring the level of speech signal clipping when the initial undistorted signal is unknown, and the parameters of the analyzed signal (sampling frequency, mean value, power, etc.) vary within wide ranges. In [4] poor experimental databases were used and the study did not investigate the parameters of the method and its influence on the performance of the method.

This paper is organized as follows. Section II presents the known methods of clipped signal detection. Sections III and IV describe a new method. Experimental studies are presented in Sections V and VI. Discussion and Conclusions are provided in Section VII.

II. Known Methods for Clipped Signal Detection

Approaches in which the source signal was unknown are studied in [5], [6]. It is understandable that to detect the clipped fragment of a speech signal, one must first evaluate the clipped level and, second, compare the value obtained to a threshold level.

In [5], a method is proposed for clipping level estimation and clipping detection based on "weighted differentiation."

S. V. Aleinik is with the Scientific department in Speech Technology Center, 4 Krasutskogo street, St. Petersburg, 196084, Russia (phone: +7-921-5830168; e-mail: aleinik@speechpro.com).

Yuri Matveev is with the University ITMO, St. Petersburg, Russia. (phone: +7-952-3669937; e-mail: matveev@mail.ifmo.ru).

This work was partially financially supported by Government of Russian Federation, Grant 074-U01.

Indeed, if the adjacent clipped signal samples $x(i)$ and $x(i-1)$ are equal (or almost equal), the value of $d(i) = x(i) - x(i-1)$ is equal to or close to zero. Accordingly, the average absolute value of $d(i)$: $D = \langle | d(i) | \rangle$ can serve as an indicator of the level of clipping: the stronger the clipped signal, the closer to zero the value of D. This method works well for slowly varying nonoscillating signals. These are the same types of signals that were examined in [5]. Unfortunately, the above condition is not fulfilled for speech signals. Voiced and unvoiced sounds in speech signals contain rapidly changing components, so even in strong clipped fragments, neighboring signal samples may be very different from one another. Thus, the value of D might not approach zero even in strongly clipped speech signals; in contrast, D fluctuates considerably and the accuracy of the method in [5] is low.

In [6] a histogram method is used to estimate the level of signal clipping: i.e. the histogram of the signal being processed is constructed and analyzed. This method in our opinion is more suitable for clipping detection. Indeed, it is known that the distribution density of the amplitudes of nonclipped speech signal can be well approximated by symmetric distributions such as the Gamma or the Laplace distribution [7]. The general shape of these distributions is a single mode curve with a smoothly decaying tail. Quite a different curve is observed in the case of a clipped signal (Fig. 2).

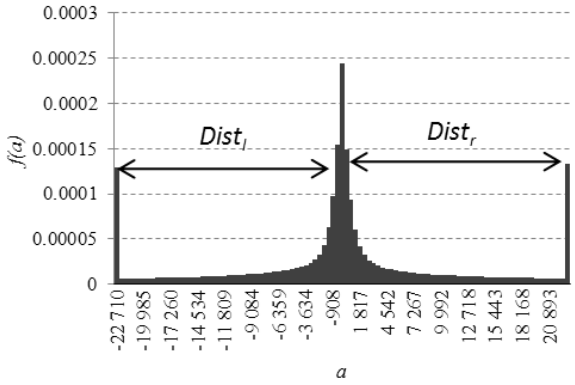

Fig. 2 Histogram of a clipped speech signal

The central part of the histogram still looks like a Laplace distribution (or symmetrical Gamma distribution). On the other hand, there are sharp bursts on the left and right tail ends, caused by the concentration of samples at the maximum and minimum values of the dynamic range.

In [6], two ways of evaluating clipping levels are proposed. The first is based on the calculation of the deviation $d_H(a)$ between the estimated normalized histogram $H_{est}(a)$ and some previously known "base" density distribution $H_{base}(a)$:

$$d_H(a) = H_{base}(a) - H_{est}(a), \quad (3)$$

where a is the abscissa of the histograms. To calculate the final clipping level, either the momentary value of $d_H(a)$ is used, or the averaged $d_H(a)$ is used over the selected parts of the histogram (excluding the central part). The second method is based on counting the number of local maxima in the tails of the calculated histogram $H_{est}(a)$.

Both methods have significant drawbacks. First, the resulting estimated value has low precision on a limited set of data. Consider, for example, frame-wise processing of a speech signal with a sampling frequency of 8 kHz, a frame length of 0.5 seconds, and with histogram bins K equal to 101. In this case the number of signal samples for the histogram estimation is N =4000. It is known that K = 101 is too high for N =4000 for statistically significant histogram estimation [8]. In this case histogram samples have high fluctuations and there are a lot of zero bins in the histogram. On the other hand, using Rice ($K = 2N^{1/3}$) or Sturges' ($K = 1 + \log_2(N)$) rules [8], we get small K (32 and 13, respectively), and, correspondingly, a high width of histogram bins. The result is the decreasing of the bursts at the tails of the histogram. In both cases counting the number of local maxima and deviation of $d_H(a)$ often provide invalid results. Second, in (3) it is necessary to know the base distribution $H_{base}(a)$.

III. The Proposed Method

A. Basics

The proposed method is also based on the analysis of the signal histogram. However, in contrast to [6], with the aim of improving performance, "base" density or histogram amplitude values are not used.

The method is based on the following considerations:

1) If the speech signal is not clipped, the tails of the histogram decrease from median point to the edges.
2) If the speech signal is clipped, there are two bursts (double-sided clipping) or one burst (single-sided clipping) at the end of the left and/or right tail(s) of the histogram.

B. Detailed Description

Consider Fig. 2. Suppose we found the very left sample of the histogram and measure its x_0 and y_0 coordinates. Then, moving along the X-axis to the histogram median, we get the current (x_i, y_i) histogram sample and compare y_0 and y_i values. If $y_0 > y_i$ we go to the next $i+1$ step; but if $y_0 \le y_i$ we do the following:

1) Calculate and store X-axis distance: $d_m = x_i - x_0$.
2) Set $x_0 = x_i$; $y_0 = y_i$ and $m = m + 1$.
3) Continue searching (do next $i+1$ step).

Thus, when we reach the histogram median, we get a set of distances: d_m, $m = 0, M$. Maximum:

$$Dist_l = \max\{d_m\},\ \mathrm{m} = 0, M, \quad (4)$$

is our intermediate result (see Fig. 2).

By analogy with the procedure described above (but starting with the rightmost sample of the histogram and decreasing the index i), we get the value $Dist_r$ (Fig. 2). Value:

$$R_{cl} = 2\max\{Dist_l, Dist_r\}/(\mathrm{X}_{\max} - \mathrm{X}_{\min}) \quad (5)$$

is the proposed clipping coefficient.

It is clear from our experiments that if signals are not clipped, histogram amplitude fluctuations are small and histogram tails rise smoothly from the right and left edges of the histogram to the median. In such cases, we often met the condition $y_0 \le y_i$; thus, the values $Dist_l$ and $Dist_r$ were small and, correspondingly, R_{cl} was also small. In contrast, when the signal is strongly clipped and there is a burst(s) at the end of the histogram tail(s), the condition $y_0 \le y_i$ will be true only near the histogram median (and sometimes never – for instance, when median amplitude is lower than the burst amplitude). Thus, the corresponding value $Dist_l$ or $Dist_r$ will be high and R_{cl} will also be high (this case is depicted in Fig. 2).

We note here that:

1) We use "max" in (5) instead of mean value with the aim of detecting single-sided clipping (because in this case one of the values, $Dist_l$ or $Dist_r$, is high, but the other is small).
2) Starting (left) and end (right) bins of the histogram may be equal to zero. These bins should be excluded from the calculation.
3) It is not necessary to calculate X-axis distances in real values for d_k calculation. It is sufficient to use the indexes of the histogram bins. In this case the denominator in (5) will be equal to the total number of histogram bins.
4) Instead of providing the sequential calculation of $Dist_l$ or $Dist_r$, it is faster to calculate a global single value $D_{\max}$ directly, increasing the left and decreasing the right current histogram indexes until they are equal, as shown below.

C. Algorithms

In the following clipping coefficient calculation algorithm, we omit the typical error check (e.g. "all bins of the histogram are equal to zero," "too few non-zero bins in the histogram," etc.).

Algorithm 1. Clipping coefficient calculation

- Calculate the histogram of the signal: $H(k), \mathrm{k} = 0, K-1$
- Find the very left k_l and the very right k_r non-zero histogram bin indexes.
- Calculate $Denom = k_r - k_l$;
- Set $y_{l0} = H(k_l)$; $y_{r0} = H(k_r)$; $d_l = d_r = 0$; $D_{\max} = 0$;
- **While** ($k_r > k_l$) **do**:
  - Increase: $k_l = k_l + 1$;
  - Decrease: $k_r = k_r - 1$;
  - If($H(k_l) \le y_{l0}$) Then: { $d_l = d_l + 1$; }
    Else: { $y_{l0} = H(k_l)$; $d_l = 0$; }
  - If($H(k_r) \le y_{r0}$) Then: { $d_r = d_r + 1$; }
    Else: { $y_{r0} = H(k_r)$; $d_r = 0$; }
  - $D_{\max} = \max\{D_{\max}, d_l, d_r\}$;
- **End while**
- Calculate clipping coefficient $R_{cl} = 2D_{\max}/Denom$;

For histogram calculation, the following well-known algorithm was used:

Algorithm 2. Histogram calculation

Let $x(n), n = 0, N-1$ is discrete time signal and K is number of bins in histogram, so:

- For all $n = 0, N-1$, find minimal $x_{\min}$ and maximal $x_{\max}$ signal values.
- Set all histogram bins to zero: $H(k) = 0, k = 0, K-1$
- **For all** $n = 0, N-1$ **do**:
  - Calculate value: $y(n) = (x(n) - x_{\min})/(x_{\max} - x_{\min})$;
  - Calculate bin index: $k = (\mathrm{int})\{Ky(n)\}$;
  - Increase histogram bin:
    If($k < N$) Then: $H(k) = H(k) + 1$;
    Else: $H(k-1) = H(k-1) + 1$;
- **End do**

Note that, we do not provide histogram normalization because it is unnecessary for the proposed algorithm.

IV. Properties of the Proposed Clipping Coefficient

Theoretical and modeling results show that:

1) $0 \le R_{cl} \le 1$ (since $D_{\max}$ cannot be more than $Denom/2$).
2) R_{cl} is invariant to the signal sampling frequency.
3) R_{cl} is invariant to the signal power and mean value (the dependence is eliminated in the histogram calculation).

We also should also point out that simple signals, such as single harmonic signals or sequences of rectangular pulses, produce $R_{cl} = 1$ even without being clipped. However, this is not a critical drawback in our opinion, because the presence of harmonics or pulses in speech overwhelmingly means that we have collided with interfering noises, which must be rejected.

V. Experimental Studies of the Proposed Algorithm

We conducted a number of experiments with different signals and algorithm parameters. The speech signals were taken from the well-known TIMIT database (16 kHz, WAV PCM 16-bits, mono). In order to avoid influence of low level noise, we deleted long pauses (longer than 0.2ms) with averaged absolute value of a signal less than 200. We then clipped the processed signals and calculated the distribution density of the coefficient R_{cl}. To produce a smooth curves, two-pass "to and fro" exponential smoothing with $\alpha = 0.8$ was used [9]. We provided hard clipping of the signals using the so-called "clipping percent" parameter $0 \le Cl_p < 100$ as follows:

Algorithm 3. Signal clipping

Let $x(n), n = 0, N-1$ be the initial discrete time signal, and $x_{cl}(n), n = 0, N-1$ be output clipped signal, so:

- For all $n = 0, N-1$, find the maximum of absolute value of the initial signal: $x_{abs,\max}$.
- Calculate threshold: $Tr = x_{abs,\max}(1 - Cl_p/100)$;
- **For all** $n = 0, N-1$ **do**:
  - If($x(n) > Tr$) Then: { $x_{cl}(n) = Tr$; }
    Else: {
    If($x(n) < -Tr$) Then: { $x_{cl}(n) = -Tr$; }
    Else: { $x_{cl}(n) = x(n)$; }
    }
- **End do**

It is clear that we have a nonclipped signal when $Cl_p = 0$ and the closer Cl_p is to 100, the stronger the signal is clipped.

The main parameters of the proposed algorithm relate to the histogram calculation, i.e. frame length N and the number of histogram bins K. It is known that the higher N is, the more accurate the estimated histogram will be, in cases when the parameter K has lower and upper bounds [10], [11]. On the other hand, our main task was not to find the optimal K for a given N as in [10], [11]. We tried to find minimal N and the corresponding K which would provide an appropriate division into the following classes: "clipped speech" and "nonclipped speech." Our experiments demonstrated that the proposed algorithm is more robust to the N and K parameters than the algorithms described above. It has been discovered that setting coarse boundaries, $N \ge 4000$ and $K \ge 101$ demonstrates good performance for the new algorithm.

For example, consider Fig. 3, which depicts distribution densities of coefficient R_{cl} for nonclipped speech signals for different N and K.

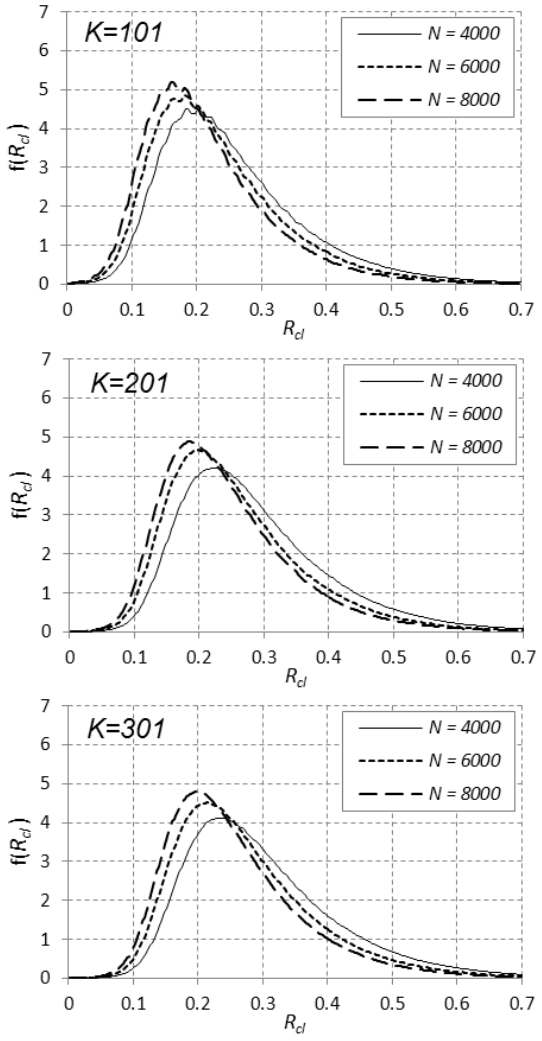

Fig. 3 R_{cl} distributions for nonclipped speech signals

Despite slight mean value shifts, all the curves are almost identical: single-mode distributions with modes in the interval [0.1, 0.2], generally shaped like a gamma distribution. An important result is that the right tail of the distributions in all cases is almost reduced to zero at $R_{cl} > 0.7$.

Fig. 4 shows the distribution densities of coefficient R_{cl} for 25% clipped speech signal for different N and K. It is seen that even for a low (25 %) level of clipping, R_{cl} can often be high, i.e. approaching 1. Increasing the parameter K causes a notable shift in R_{cl} distribution of clipped speech to the right.

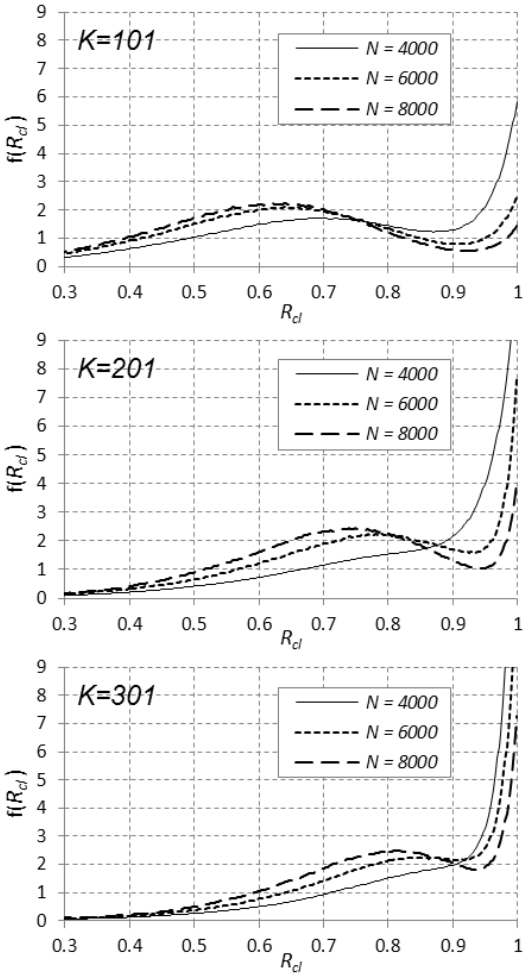

Fig. 4 R_{cl} distributions for 25 % clipped speech signal

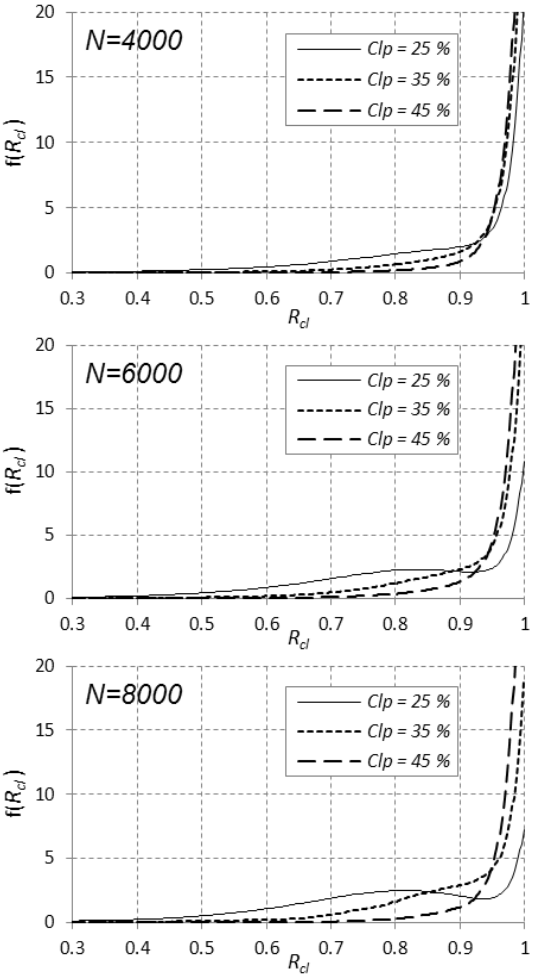

Fig. 5 R_{cl} distributions for K = 301 and different N and Cl_p

Fig. 5 demonstrates the distribution densities of coefficient R_{cl} for fixed K = 301 and for a different N and clipping percentage Cl_p. For the given N and K, increasing Cl_p leads to a strong shift in the corresponding distribution to the right: if $Cl_p = 45\%$, values of $R_{cl} > 0.9$ were found in more than 90% cases for all N.

The curves for nonclipped and Cl_p = 25 % clipped speech and K = 301 are depicted in Fig. 6.

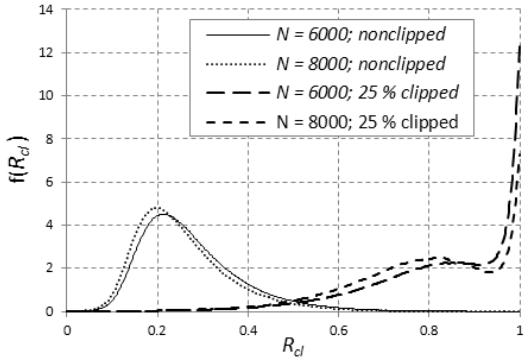

Fig. 6 R_{cl} distributions for K = 301, different N and nonclipped and 25 % clipped speech signal

Corresponding probabilities of False Alarm (FA) and Right Detection (RD) for the given Cl_p, N and K are presented in Table I below.

TABLE I
FALSE ALARM AND RIGHT DETECTION

R_{cl}	FA 6000	RD 6000	FA 8000	RD 8000
0.30	0.330	0.993	0.272	0.992
0.40	0.125	0.981	0.096	0.978
0.50	0.044	0.955	0.032	0.945
0.60	0.013	0.899	0.010	0.869
0.70	0.004	0.792	0.003	0.725

We therefore conclude that the threshold level for R_{cl} in the interval [0.4, 0.6] is acceptable for adequate separation of clipped / nonclipped speech for Cl_p > 25 % for the following algorithm parameters: N = 6000, 8000 and K = 301.

VI. Example of Real Speech Signal Processing

Figs. 7-9 depict the processing results for the TIMIT database phrase "Tim takes Sheila to see movies twice a week" (N = 6000, K = 301) for different Cl_p.

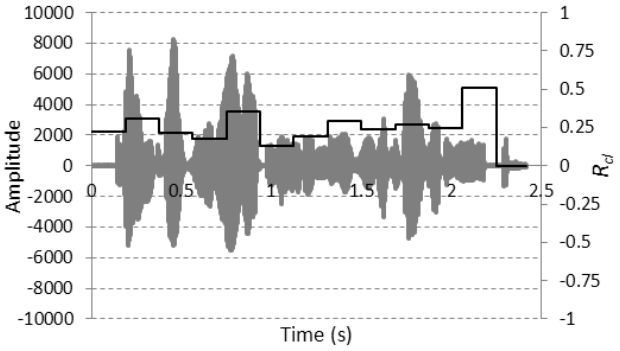

Fig. 7 Speech signal and R_{cl} for Cl_p = 0

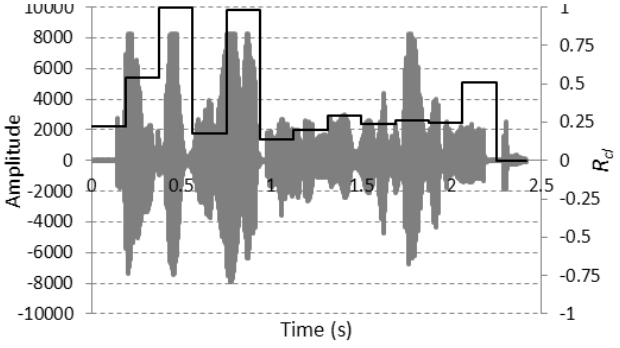

Fig. 8 Speech signal and R_{cl} for Cl_p = 25

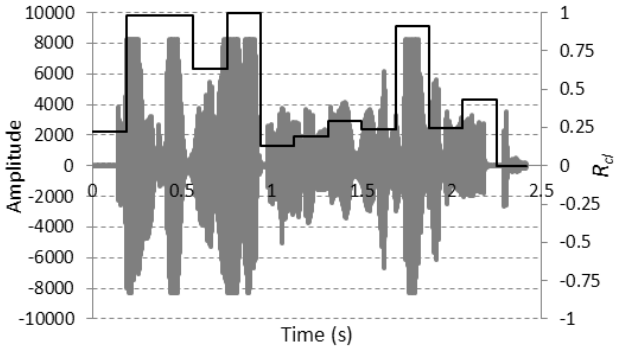

Fig. 9 Speech signal and R_{cl} for Cl_p = 50

The proposed method gives almost 100% performance on the present example, where threshold = 0.55: all clipped parts of the speech signal are detected (the corresponding R_{cl} > 0.55), while nonclipped frames are marked as nonclipped.

VII. Discussion and Conclusion

This paper has presented a method for detecting clipped fragments of a speech signal. The method is robust to the signal parameters, such as mean value, power, etc. Parameters of the method were investigated using computer simulation. Although the method shows good performance, we should point out that:

1) A sharp increase in the coefficient R_{cl} to values near 1 when Cl_p is more than 50% suggests the method is more suitable for clipping detection than for clipping level measurement (as it was claimed in [4]), especially for high clipping levels.
2) It must be remembered that on simple harmonic signals, the method gives R_{cl} = 1 even without clipping.
3) Our experiments show that the method may also give overestimated results when the analyzed frame contains few nonzero and many zero samples.

References

[1] H. Chen, A.M. Haimovich, "Iterative Estimation and Cancellation of Clipping Noise for OFDM Signals," *IEEE Commun. Lett.*, vol. 7, no. 7, pp. 305–307, 2003.
[2] S.V. Zhidkov, "Detection of Clipped Code-Division Multiplexed Signals," *Electronics Letters*, vol. 41, no. 25, pp. 1383-1384, 2005.
[3] J. Kim, "Method and Apparatus for Evaluating Audio Distortion," US Patent 005402495, Int.Cl. H04B 15/00, 1995.
[4] S. Aleinik, Yu. Matveev, A. Raev, "Method of Evaluation of Speech signal clipping level," *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, vol. 79, no. 3, pp. 79-83, 2012, (in Russian).
[5] T.E. Riemer, M.S. Weiss, M.W. Losh, "Discrete Clipping Detection by Use of a Signal Matched Exponentially Weighted Differentiator," in *Proc. of the IEEE Southeastcon*, New Orleans, Louisiana, 1990, pp. 245–248.
[6] T. Otani, M. Tanaka, Y. Ota, S. Ito, "Clipping Detection Device and Method." US Patent 20100030555 A1, Int.Cl. G10L 21/02, 2010.
[7] L. R. Rabiner, R. W. Schafer, *Introduction to Digital Speech Processing.* Now Publishers Inc., Hanover, MA, 2007.
[8] D. M. Lane (editor), *Introduction to Statistics*, Rice University, Online Edition, http://onlinestatbook.com/Online_Statistics_Education.pdf.

[9] P. Ignatov, M. Stolbov, S. Aleynik, "Semi-Automated Technique for Noisy Recording Enhancement Using an Independent Reference Recording," in *Proc. 46th International Conference of the Audio Engineering Society*. Denver, USA, pp.57-64, 2012.
[10] K.H. Knuth, "Optimal Data-Based Binning for Histograms," 2006. arXiv:physics/0605197, http://arxiv.org/pdf/physics/0605197.pdf.
[11] Wand M. P., "Data-Based Choice of Histogram Bin Width," *The American Statistician*, vol. 51, No. 1, 1997, pp. 59-64.