ARCHIVES OF ACOUSTICS
32, 4 (Supplement), 165–171 (2007)

DETECTION AND REMOVAL OF "SMACKING" ARTEFACTS FROM LECTOR SPEECH RECORDS

Marek JASKUŁA, Hanna PERUŻYŃSKA

Szczecin University of Technology
Faculty of Electrical Engineering
26 Kwietnia 10, 71-126 Szczecin, Poland
e-mail: Marek.Jaskula@ps.pl

(*received July 15, 2007; accepted October 17, 2007*)

In the paper, analysis of "smacking" artefacts, introduced to the speech signal by a lector, was shown. A working algorithm to reduce amount of smacks in the waveform was presented. The algorithm performs detection of distortions in the time domain using differentiation of the signal. The removal routine is based on Dicrete Wavelet Transform (DWT) and Inverse Discrete Wavelet Transform (IDWT).

Keywords: speech enhancement, wavelet transform, impulse noise.

1. Introduction

Noise was always an important issue in sound acquisition systems. It happens rarely that the waveform is not corrupted with some kinds of artefacts. When caused by recording medium or devices working in the neighborhood, they can be easily avoided by changing the equipment or surrounding. More complicated task is to eliminate artefacts introduced to the signal by the speaker. It is not always possible to change a person who gives a speech. Furthermore, while there is a wide range of tools specialized to exclude noise generated by different kinds of machines, there are no similar solutions for artefacts produced by a man. In this paper we demonstrate a working algorithm to reduce amount of "smacking" artefacts in the speech tract. This particular type of distortion arise mainly from saliva snapping in the speaker's mouth. It is not only decreasing the quality of the signal but can also annoy the listener, making hard to focus on the record's content.

Currently, processing of the records degradated by "smacking" artefacts is made manually in sound editors. The operator locates smacks by first listening to the audio segment and then zooming in on the suspected area. Then the smack is usually replaced with a tiny bit of the waveform that comes immediately before or after it. This is a time-consuming process, mainly because distortions tend to occur in numerous groups at random positions in the waveform. The job also requires a lot of experience as the artefacts frequently have low amplitude and very short duration. Even highly qualified editor can omit some smaller but audible defects.

2. Materials and methods

Two different speech tracks, deeply degradated by "smacking" artefacts, were analyzed [2]:

- 2 minutes record of a woman speaking in Polish,
- over 20 minutes record of a man speaking in English.

Both records have ca. 600 artefacts with duration ranging from less than 0.2 ms up to 7 ms in extreme cases.

Single distortion, perceived by a listener as a smack, is generally cluster of pulses with amplitude several times higher than the amplitude of surrounding sound (see Fig. 1). Artefacts often consist of a relatively short sharp initial pulse followed by decaying high-frequency oscillations, although small oscillations can as well precede the initial pulses. Smacks can also be formed by few pulses of similar amplitude.

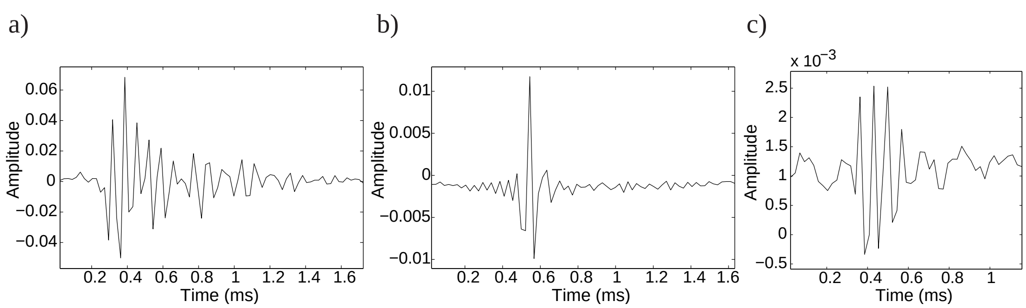

Fig. 1. Different types of smacks observed in analyzed records.

The duration of initial pulses varies from 0.07 ms to 0.14 ms. They have flat spectrum in wide range of frequencies which makes smacks easily noticeable in the spectrogram of a section of speech. As can be seen in Fig. 2 it is not possible to define one range of frequencies common for all distortions, so simple bandpass filtration can not eliminate them effectively. Considered artefacts can be compared to impulsive noise which originates from variety of sources such as clicks from computer keyboards or physical defects in the recording medium. Figure 3 shows scratch pulse from a damaged gramophone record. Common feature of scratches and smacks is the presence of short duration initial pulse, therefore tools optimized for scratches elimination can be helpful in the process of smacks detection. Unfortunately they are not suitable for smacks removal, due to the different nature of decaying oscillations.

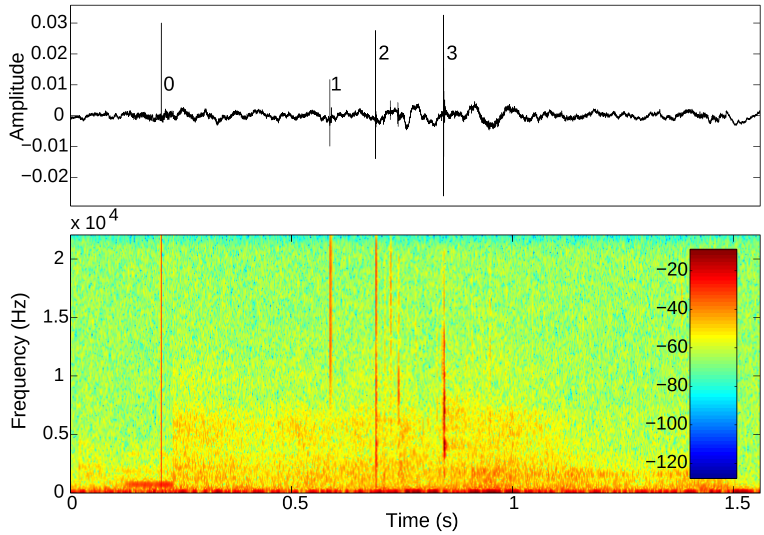

Fig. 2. Waveform distorted with smacks and its spectrogram, 0 – pulse of 1-sample duration, numbers 1, 2, 3 – smacks.

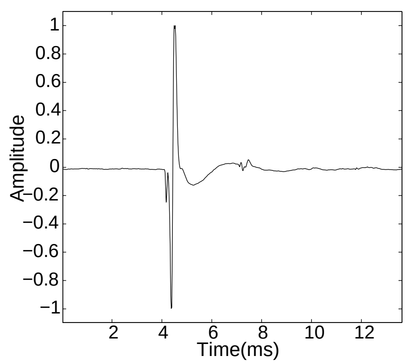

Fig. 3. Scratch pulse from a damaged gramophone record.

3. Elimination scheme

Developed algorithm, fitted for eliminating of smacks from speech tracks, was implemented in Matlab program. It consists of two basic steps (see Fig. 4):

- signal analysis – detection of smacks,
- signal modification – removal of smacks.

The analysis stage, in which detection of distortions is performed, starts with detection of smacks beginnings. Then the endings of detected defects are established and the points in which signal crosses zero amplitude are searched for (left from smack beginning and right from its ending). Time intervals marked in detection process are further

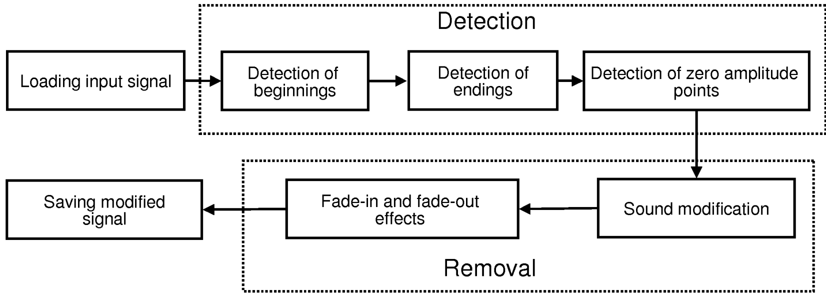

Fig. 4. Elimination scheme.

modified in order to remove smacks. To minimize leaping of the amplitude at the ends of the intervals fade-in and fade-out effects are introduced. They can work properly thanks to the fact that intervals begin and end in points of zero amplitude.

4. "Smacking" artefacts detection

Smacks are normally more distinct and detectable in the time domain than in the frequency domain, therefore developed algorithm uses time-domain signal processing for their detection.

The basic idea is to find short fragments of the track with amplitude sufficiently high in comparison to the surrounding sound. Detection is improved by using differentiation. Differencing utilizes characteristic differences between the noise and the signal. It emphasizes sudden, fast changes of the signal, therefore smacks become more detectable (see Fig. 5).

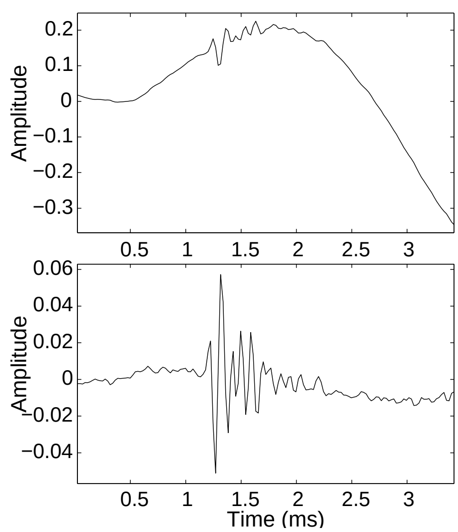

Fig. 5. Signal distorted with smack and its derivative.

Figure 6 shows the way in which beginning of each smack is established. The user determines the duration of searched artefacts (`Duration`) and the ratio of their amplitude to amplitude of surrounding sound (`Ratio`). Increasing the ratio and decreasing the duration we can achieve better reliability of the system. The length of the portion of the signal which is used for comparison can be changed as well (`Portion`). Algorithm calculates mean values of amplitude before and after each sample of the sound (`Before`, `After`) taking predetermined smack duration into account. Next value of analyzed sample (`i`) is compared with computed values. If it is sufficiently high, number of the sample is remembered as the beginning of the smack (`SmackBeg`). Usually it is the location of the highest pulse in the smack. After defining the beginning, the ending of the smack is searched for. Figure 7 shows the way in which it is done. The algorithm calculates the mean amplitude before detected beginning (`BeforeBeg`) with the shift backwards, in case there are some small oscillations before the initial pulse. The shift (`Shift1`) and the number of utilized samples (`Shift2`) is defined by the user. For each sample, forwards from the detected beginning of the smack, mean amplitude is computed from the same number of samples as before (`AfterBeg`). If the computed values are similar, analyzed sample is remembered as the end of the smack (`SmackEnd`). The user has influence on the similarity of the computed mean values (`EndCoef`).

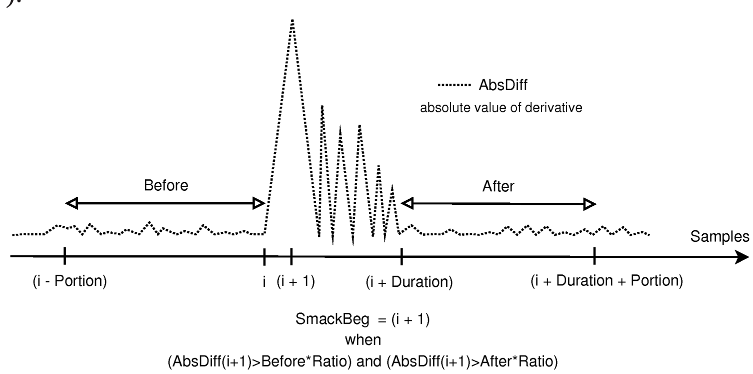

Fig. 6. Detection of smacks beginnings.

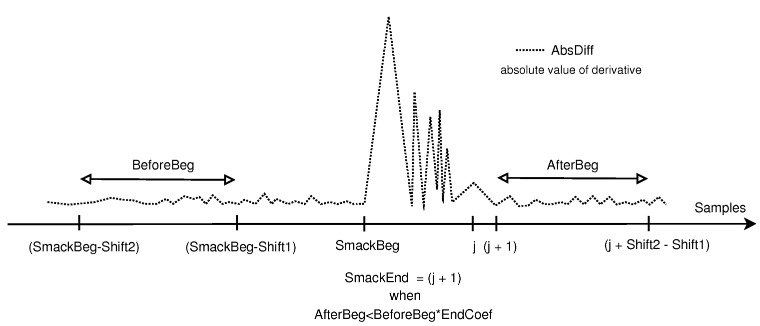

Fig. 7. Detection of smacks endings.

After the limits of the smack are detected the algorithm is looking for the points in which signal crosses zero amplitude. It is necessary for the correctness of fade-in and fade-out effects introduced in the next stage.

5. “Smacking” artefacts removal

To make the process of artefatcs elimination fully automatic, the algorithm of detection is combined with the algorithm of artefacts removal. The distorted fragments of the soundtrack, marked in the detection stage, are further modified in order to remove artefacts. This process is based on Discrete Wavelet Transform (DWT) and Inverse Discrete Wavelet Transform (IDWT). The signal is decomposed with DWT into five groups of coefficients. Four of them contain information about smacks. The inverse transform is performed on the group that is free from distortions, earlier the other coefficients are set to zero. As a result of this action undistorted waveform is achieved.

During described process other useful, high-frequency content is removed from the signal. Although it has very low amplitude, lack of it can be noticed by a careful listener. Therefore the algorithm adds discarded information to the modified piece of sound. It is obtained from the waveform that comes immediately before the disturbed piece of soundtrack.

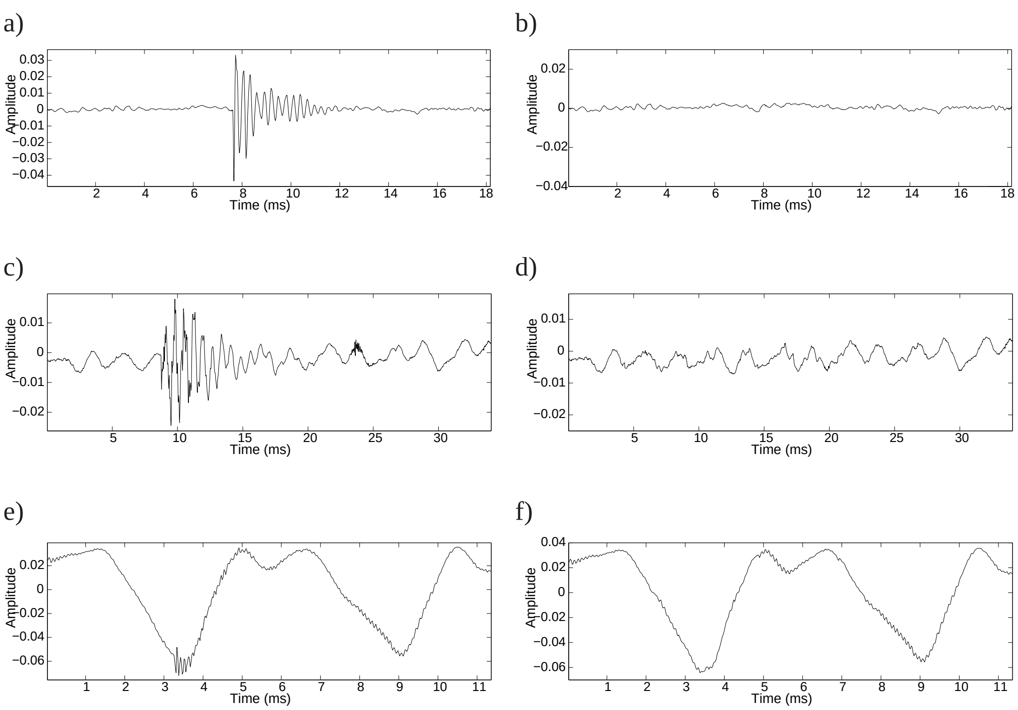

Fig. 8. Several results of described signal processing (a, c, e – original signals; b, d, f – after processing (respectively)).

At the end, the borders of modified intervals are smoothed with fade-in and fade-out effects. Fade-out is introduced before and fade-in after each of the two breakpoints. What we attain is smooth transition between unaffected and changed part of the waveform. Figure 8 shows several results of the described signal processing.

6. Conclusions

Presented algorithm was implemented as a Matlab function. It enables the user to change parameters of artefacts detection and removal, to adjust this routines to suit smacks of different amplitude and duration. It also draws the effects of signal processing, writes modified track as an audio file and creates reports containing information about the processed signal and record of its work.

Quality of distorted speech tracks processed with the program was improved. The number of smacks in both of them decreased noticeable. Corrupted signals required more than one processing with different setting of the parameters. The easiest and safest way to upgrade quality of the speech track is to remove first very short then longer artefacts. To avoid signal damaging it is better to run the algorithm on a small piece of sound before processing the entire data.

References

[1] Godsill S. J., Rayner P. W., *Digital Audio Restoration. A Statistical Model Based Approach*, Springer – Verlag London Ltd, London 1998.

[2] Perużyńska H., *Analiza i przetwarzanie wybranych typów artefaktów dźwiękowych z wykorzystaniem programu Matlab i języka Nyquist*, Master Thesis supervised by Marek Jaskuła, Szczecin University of Technology, 2007.

[3] Vaseghi S. V., *Advanced Digital Signal Processing and Noise Reduction*, John Wiley & Sons, Ltd., Chichester 2000.