Digital Signal Processing 46 (2015) 253–265

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Localization of impulsive disturbances in audio signals using template matching ☆

Maciej Niedźwiecki, Marcin Ciołek *

Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, Poland

A R T I C L E I N F O

Article history:
Available online 5 August 2015

Keywords:
Outlier detection and elimination
Adaptive signal processing
Audio restoration
Digital archives

A B S T R A C T

In this paper, a new solution to the problem of elimination of impulsive disturbances from audio signals, based on the matched filtering technique, is proposed. The new approach stems from the observation that a large proportion of noise pulses corrupting audio recordings have highly repetitive shapes that match several typical "patterns". In many cases a representative set of exemplary pulse waveforms can be extracted from the episodes of silence preceding and succeeding the recorded audio material. Based on such a set, a relatively small number of typical noise patterns, called click templates, can be established. To localize noise pulses, the appropriately modified click templates can be correlated with the sequence of one-step-ahead prediction errors yielded by the model-based signal predictor. It is shown that template matching is an efficient and computationally affordable disturbance localization technique – when combined with the classical detection method based on autoregressive modeling, it can improve restoration results. Since click templates can be created for a particular set of recordings, obtained using a particular audio equipment, an important feature of the proposed approach is its source adaptivity. Even though the paper is focused on restoration of archive recordings, the proposed approach is useful in a much wider context, e.g., it can be applied to elimination of impulsive disturbances corrupting telecommunication channels.

1. Introduction

Archived audio recordings are often degraded by impulsive disturbances and wideband noise [1,2]. Clicks, pops, ticks, crackles and record scratches are caused by aging and/or mishandling of the surface of gramophone records (shellac or vinyl), specks of dust and dirt, faults in the record stamping process (e.g. gas bubbles), and slight imperfections in the record playing surface due to the use of coarse grain filters in the record composition. In the case of magnetic tape recordings, impulsive disturbances can be usually attributed to transmission or equipment artifacts (e.g. electric or magnetic pulses).

Wideband background noise, such as the so-called surface noise of magnetic tapes and phonograph records, is an inherent component of all analog recordings.

Elimination of both types of disturbances from archive audio documents is an important element of saving our cultural heritage. The Polish Radio Archives and the Polish National Library Archives alone contain more than one million archive audio documents with different content (historic speeches, interviews, concerts, studio music recordings etc.), saved on different media, such as piano rolls, phonograph and gramophone records, magnetic tapes etc. The British Library Sound Archive (which is among the largest collections of recorded sound in the world) holds over three million recordings, including over one million of disks and 200,000 tapes. Digitization of these documents is an ongoing process (in Poland carried out, among others, by the Polish National Digital Archives), which will be very soon followed by the next, obvious step – audio restoration. This makes research on audio restoration technology both practically useful and timely.

The majority of known approaches to elimination of impulsive disturbances from archive audio signals are based on adaptive prediction – the autoregressive (AR) or autoregressive moving average (ARMA) model of the analyzed signal is continuously updated and used to predict consecutive signal samples [3–14]. In the simplest case, further referred to as basic detection scheme, a "detection alarm" is raised, and the predicted sample is scheduled for re-

☆ This work was supported by the National Science Centre under the agreement UMO-2013/09/B/ST7/01582.

* Corresponding author.
E-mail addresses: maciekn@eti.pg.gda.pl (M. Niedźwiecki), marcin.ciolek@pg.gda.pl (M. Ciołek).

http://dx.doi.org/10.1016/j.dsp.2015.07.009
1051-2004/

construction, whenever the absolute value of the one-step-ahead prediction error becomes too large, namely when it exceeds a prescribed multiple of its estimated standard deviation. The test is then extended to multi-step-ahead prediction errors – detection alarm is terminated when a given number of samples in a row remain sufficiently close to the predicted signal trajectory (or when the length of detection alarm reaches its maximum allowable value). Finally, once the pulse is localized, the corrupted samples are interpolated (using the same signal model which served for detection purposes) based on the uncorrupted neighboring samples. A more sophisticated, Bayesian solution to the problem of noise pulse detection and signal reconstruction (also based on AR modeling) was presented in [6] and [7]. In both cases noise pulses were modeled as additive bursts of noise.

The basic detection scheme was subject to several modifications and extensions.

In [5] and [8] the task of simultaneous signal identification, outlier detection and signal reconstruction was stated as a nonlinear filtering problem and solved using the theory of extended Kalman filter (EKF). The EKF algorithm can be viewed as a combination of two Kalman filters coupled in a nonlinear fashion – the filter designed to track time-varying parameters of the AR signal model, and another one used for the purpose of detection and reconstruction of corrupted samples. As later shown in [9], applying the certainty equivalence projection technique one can partition the EKF algorithm into two weakly coupled subalgorithms responsible for model parameter tracking and signal monitoring/reconstruction, respectively. This has two important practical implications. First, the Kalman filter based parameter tracker can be replaced with a more convenient (easier to tune) exponentially weighted least squares (EWLS) algorithm [15,16]. Second, the detection/reconstruction algorithm can be put down in the order-recursive form, which results in major computational savings.

Even though yielding satisfactory results when applied to archived music, the AR-model based reconstruction often fails on speech signals, especially those with strong voiced episodes. Since voiced speech sounds are formed by exciting the vocal tract (represented by the AR model) with a periodic train of glottal air pulses, the outlier detector can easily confuse pitch excitation with impulsive noise, which usually results in audible signal distortions. The problem mentioned above can be alleviated if the sparse autoregressive (SAR) model of the audio signal is used instead of the AR model [12]. SAR models capture both short-term correlations (formant structure) and long-term correlations (pitch structure) of the analyzed sound. Owing to this, unlike outlier detectors based on conventional AR models, detectors that incorporate SAR models usually do not confuse pitch-related pulses with noise pulses. This significantly reduces the number of false alarms.

Restoration of stereo recordings can be performed by splitting left/right audio tracks and processing them separately. However, improved results can be obtained if both tracks are modeled jointly using the vector autoregressive (VAR) or sparse vector autoregressive (SAR) modeling approach [14]. The benefits of VAR/SVAR modeling can be observed both at the outlier detection stage (more accurate localization of noise pulses) and at the signal interpolation stage (the undistorted material in one track can be used to "repair" the corrupted fragment in the other track).

When the archive audio signal is analyzed sequentially, forward in time, a sample is regarded as an outlier if it is "inconsistent" with the signal past, which is indicated by excessive values of prediction errors. When signal characteristics change abruptly, e.g. when an entirely new sound starts to build up, all causal detection schemes are prone to generate false detection alarms, calling in question uncorrupted signal samples simply because they do not match the signal past. Since such samples are consistent with the signal "future", rather than its "past", the number of false alarms can be significantly reduced if results of forward-time detection are combined with the analogous results of backward-time detection. The latter can be obtained by means of processing audio signal backward in time (provided, of course, that the entire recording is available). In addition to reducing the number and length of false alarms, bidirectional processing allows one to carve detection alarms more carefully (smaller number of overlooked noise pulses, better front/end matching of noise pulses). The set of local, case-dependent fusion rules that can be used to combine forward and backward detection alarms was proposed and experimentally verified in [13].

The common feature of the approaches summarized above is that they all incorporate outlier elimination schemes which do not rely on any information about the size and shape of noise pulses – even if such a prior knowledge is available. To the best of our knowledge, apart from the method described in [4], which focuses on very long disturbances such as record scratches, the only approach proposed so far, which incorporates prior knowledge about noise pulses into pulse detection/elimination procedure, is that described in the recent paper of Ávila and Biscainho [11]. The Bayesian pattern matching procedure proposed in [11] is based on two sequentially sampled models: the AR model of the clean audio signal (with adjustable autoregressive coefficients and adjustable driving noise variance), and an explicit model of the impulsive disturbance (exponentially decaying pulse with adjustable location and shape parameters). The problem of joint detection and estimation of corrupted samples is solved by means of Gibbs sampling – the joint posterior distribution of the clean signal, its AR-model parameters and noise pulse parameters, is searched numerically using a variant of the Metropolis–Hastings algorithm. The resulting numerical iterative procedure is computationally very demanding.

The approach described in this paper is explicit and much simpler. It originates from the observation that in many cases a representative set of impulsive disturbances can be extracted, using simple detection techniques, from the episodes of silence preceding and succeeding the recorded audio material (e.g., separating successive tracks on long-playing records). Based on such a set, a relatively small number of typical noise patterns (click templates) can be established and further used for detection purposes.

The contribution of the paper is twofold. First, we propose a new method, based on analysis of the pulse similarity graph, that allows one to create the library of click templates. Second, we show how typical noise patterns can be detected and localized in the archive recording using the matched filtering technique. We demonstrate that, when combined with the classical AR-model based detection methods, such approach can noticeably improve restoration results.

For clarity reasons, our presentation will be restricted to the sequential AR-model based noise pulse elimination scheme, which is suitable for on-line processing of music signals. Extension of the obtained results to sparse modeling and/or bidirectional processing is straightforward.

Remark. The matched filtering technique was proposed in early publications on elimination of impulsive disturbances [3,4]. The authors of the above-mentioned papers analyzed an impact that an *idealized* (Kronecker-type) noise pulse has on the output of the AR-model based inverse filter. They suggested that in order to localize such pulses in the input (corrupted audio) signal, one could convolve the sequence of one-step-ahead signal prediction errors, yielded by the inverse filter, with the sequence made up of autoregressive coefficients (put in reverse order), and threshold the obtained results. Quite clearly, this approach does not incorporate any knowledge of typical noise patterns. It can only be used to isolate short unimodal pulses. The technique described in [4] is more