Comparison of Autoregressive Model Based Perceptible Click Detection Algorithms

Arda Özdoğru[1]

[1]Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, 166 27 Praha, Czech Republic

ozdogard@fel.cvut.cz

Abstract. *In this paper, Autoregressive (AR) model based click detection algorithms are compared in terms of their ability to detect only the perceptible impulsive noise (clicks) contained in the audio signals taken from damaged vinyl records. All the algorithms are using the same principle of detection but they are all improved differently in the mentioned literature. The test audio signals are previously classified to be containing or not containing clicks. Comparison criterion is based on a previously introduced custom function, defined through the correct and the false detection ratios. Run time comparisons of the algorithms are also made. Results showed that a Simple AR model algorithm performed the best, both in detection and in run time, amongst the compared AR algorithms.*

Keywords

Autoregressive, Perceptible, Click Detection, Model Comparison.

1. Introduction

Impulsive noise detection is a widely researched area with many applications, especially in audio and speech processing. Although there are various methods for such detection in audio signals [1, 2, 3], most commonly used algorithms for detection of impulsive noises (from here on referred as clicks) in audio signals include Autoregressive (AR) models.

Many algorithms are focused on the restoration of the clicks after the detection occurs [4, 3], even for the clicks that are not perceptible which can cause distortion. For some applications the restoration is not needed. An example to that can be made as the necessity of detecting clicks caused by faulty manufacturing of vinyl records for further analysis.

Evaluation of perceptible click detection algorithms using different models is done on the samples taken from damaged vinyl records at [5]. The samples contain various genres of music with vocal, rhythmic and harmonic elements. The aim of this paper is to test and to compare various AR models on their ability to correctly detect the perceptible clicks for the audio samples, which are categorized by averaging results of listening tests with a group of people, explained in detail in [5].

During the evaluation, the algorithms that have a separable detection and restoration sections are taken into consideration and the comparison was based on the output of the detection sections. The AR-based models under investigation are as follows: [6] causal and semi-causal with open or closed loop detection schemes utilized in uni-directional or bi-directional processing, [7] AR model improved by fusion parameter; [8] simple AR model, [9] matched filter based AR model. The results for the last two algorithms are directly taken from [5] since the evaluation criterion is the same.

2. Autoregressive Models

In an audio signal, corrupted sequence of N samples can be represented as:

$$y(n) = s[n] + i[n]c[n], \qquad n = 0, ..., N-1 \tag{1}$$

In this generalized formula of audio degradation, it is assumed that the noise samples $c[n]$, which are regularly between 1 and 50 samples, are uncorrelated with the audio signal samples $s[n]$ and $i[n]$ denotes the *'switching'* of the noise samples where $i[n] = 1$ when the click is present and $i[n] = 0$ otherwise. For the algorithms, the main purpose is to detect the starting and ending of *'switching'*, which corresponds to the presence of a click [10]. Statistics of $c[n]$ determine the amplitude characteristics of the corrupting process [10].

Audio signals of length N can be represented as the AR model with p^{th} order

$$s[n] = \sum_{k=1}^{p} a[k]s[n-k] + e[k], \qquad n = p, ..., N-1 \tag{2}$$

where $a[k]$ denotes the k^{th} AR coefficient and the $e[k]$, which has a Gaussian distribution, denotes the prediction error. When there is a click present after the availability of uncorrupted signals, the error $e[k]$ is likely to be very high compared to the steps that doesn't contain corrupted samples [5].

In the simplest form, $e[k]$ can be compared with a threshold to detect the boundaries of the clicks. Since the investigation characteristic changes over time, the estimation of the AR parameters are made by solving Yule-Walker equations for blocks of samples

$$[1, -\hat{a}_1(t), ..., -\hat{a}_n(t)]\mathbf{R}(t) = [\hat{\rho}(t), 0, ..., 0] \quad (3)$$

$$\mathbf{R}(t) = \begin{bmatrix} r_0(t) & \cdots & r_n(t) \\ \vdots & \ddots & \vdots \\ r_n(t) & \cdots & r_0(t) \end{bmatrix} \quad (4)$$

where

$$r_i(t) = (1/L)p_i(t), \quad i = 0, 1, ..., n \quad (5)$$

can be considered as the local (block-wise) estimate of the i^{th} autocorrelation coefficient of $y(t)$, which can be found using the equation

$$p_i(t) = \sum_{l=k}^{k} w(l)w(l-i)h_i(t-k+l), \quad (6)$$

where

$$h_i(t) = y(t)y(t-i) \quad (7)$$

and $w(l)$ is bell shaped weighting function having maximum in the center and smoothly decaying to 0 at the edges and the normalizing constant is the energy of $w(l)$,[6].

In the following subsections, the differences of the investigated algorithms from the simple model are briefly mentioned. For detailed explanations, it is advised to check the corresponding literature.

2.1. Causal, semi-causal, uni/bi-directional detection

In the algorithm suggested in [6], causal detection is made by the comparison by a threshold which is at least 3σ outliers of Gaussian distributed error estimates obtained from the predictions based on the previous uncorrupted samples. On the other hand, Semi-causal detection controls the prediction and the interpolation error, which is triggered after prediction error detection, based statistics. Interpolation error depends on the previous and the future samples around the click. Interpolation error usually gets much higher values compared to prediction error, hence a second threshold is adapted for it. Open-loop, which detects the corrupted samples as a whole, and closed-loop, which detects and classifies samples one by one, versions are used unidirectionally, which sweeps through the signal $y[n]$ ones, and bidirectionally, which reverses $y[n]$ and checks for the consistency of the detection. The detection of both directions are then combined by certain rules which are described in [6], in this paper for comparison results, fusion rules based combinations are taken, resulting in total of 8 possible options that are compared.

2.2. Fusion parameter improved AR model

In the algorithm suggested in [7], single burst of click samples is detected by a simple prediction error based thresholding. However, for a frame containing multiple bursts, a *fusion parameter b* is added to the detection scheme to limit the maximum number of burst samples. Overlap-and-Add method is used to process *locally stationary* blocks of the signal.

2.3. Simple AR and Matched Filter AR model

Simple AR model, described in [5], applies inverse AR filter to the corrupted signal $y[n]$ and thresholds the prediction error $e[n]$ given by the following equation:

$$e[k] = y[k] - \sum_{k=1}^{p} \hat{a}[n]y[k-n] \quad (8)$$

For the implementation of Matched Filter, equation (8) means filtering the corrupted signal with $p+1$ length inverse filter with the coefficients $[-1, -a_1, ..., -a_p]$. This signal then can be filtered by a filter having a time-inverse impulse response of the inverse filter to further emphasize the clicks [5]. Increased sensitivity of this method causes reduced precision of localization of the clicks [10]. To increase the localization of the clicks, bidirectional approaches described in [5] are used.

3. Evaluation Method

The evaluation of the algorithms is done by, the described stimuli set in [5], a set of (90) dynamically preserved stimuli of wav files (mono, 44.1kHz, 16 bit depth) of equal length (800 ms) that are taken from damaged vinyl records. Offset and onset of the samples are shaped with a 80 ms raised cosine ramps. The stimuli contains audio data from various genres of music. The stimuli is controlled for the classification of perceptual clicks through listening tests by 17 listeners aged between 21-47 years old. The stimuli, which are voted over %75 as containing clicks and the one that are voted over %75 as not containing clicks, are used in the evaluation of the algorithms.

Comparison is made by, criteria suggested in [5], measuring the correct detection rate:

$$R_{CD} = N_{CD}/N_C, \quad (9)$$

and false detection rate:

$$R_{FD} = N_{FD}/N_{CF}, \quad (10)$$

where N_{CD}, N_C denotes the number of correct detection and the number of stimuli marked by having perceptual clicks, respectively, and N_{FD}, N_{CF} denotes the number of

false detection and the number of stimuli marked as not having perceptible clicks, respectively. The performance evaluation is done by customary criteria of

$$d' = z(R_{CD}) - z(R_{FD}), \tag{11}$$

where $z()$ is the inverse of the normal distribution function.

The algorithms described in [6], are manipulated to run only until the end of detection part in MATLAB. The computation times are also measured for the stimulation of 90 samples. To measure the computation time, the AR algorithms in [5] and [6] are run on the same computer with the following properties: Intel Core i5-8250U x64, CPU @1.60GHZ with 4 cores and 8 logical units, 8GB RAM running Microsoft Windows 11 Home. The Fusion AR results are obtained from the utilization of the online demo, which is accessible from the reference, hence no time measurements are taken.

Algorithm	$R_{CD}(\%)$	$R_{FD}(\%)$	d'	$time(s)$
Uni, Causal, Open	91.18	66.67	0.921	25.79
*Uni, Causal, Closed	100	69.23	Inf	23.76
Uni, Semi, Open	91.18	56.41	1.190	23.05
*Uni, Semi, Closed	100	61.54	Inf	23.42
Bi, Causal, Open	97.05	64.10	1.528	25.25
*Bi, Causal, Closed	100	66.67	Inf	23.24
Bi, Semi, Open	97.05	64.10	1.528	23.00
*Bi, Semi, Closed	100	66.67	Inf	23.37
Fusion AR	100	97.43	Inf	-
Simple AR	88.24	**15.38**	**2.207**	**1.19**
Matched AR	82.35	23.08	1.665	5.76

Tab. 1. Comparison Results

4. Results and Discussion

The best performing algorithms are highlighted in the table in their respective best performing areas.

The ones mark with "*", which are part of the algorithms described in [6], can not be considered as well performing algorithms in terms of *perceptual* click detection because both the correct detection and false detection rates are high, even though the prediction thresholds were tried up to $\mu_\alpha = 8.0$ (maximum mentioned in [6] is ($\mu_\alpha = 4.5$) and the interpolation thresholds were tried up to $\mu_\beta = 8.5$ (maximum mentioned in [6] is $\mu_\beta = 4.5$). It is said in [6] that as the threshold is increased the detection gets duller. However, when the thresholds are further increased, the click detection was not proper. Normally, with high thresholds were said to miss out on small clicks, still the false detection was high for the real records taken from damaged vinyls.

The algorithm Fusion AR, described in [7], is noted to run optimally with the default settings on the online demo ($k = 2, b = 20, p = 302, nwindow = 2416$). However, it was observed that very high number of clicks are detected (in average 272 clicks/sample). Second trial with only detection threshold increased to the allowed maximum value ($k = 3$) has been done. Although the average detection has dropped to 60 clicks/sample, still, in all the samples clicks (with shorter average lengths) are detected except one which was marked as click-free and the algorithm detected zero clicks. It can be said that this algorithm is, also, too sensitive for *perceptual* click detection.

Simple AR has shown the best results in the compared algorithms in terms of false detection ratio and the d' criteria. Although, the correct detection ratio is lower compared to other algorithms except the Matched AR, Simple AR has the lowest false detection ratio which results in the best d' score. This method has also proved to be a faster algorithm compared to others, running in 1.19 seconds because of its simple nature. Matched AR has the second best score which is slightly higher than the bidirectional open-loop causal and semi-causal algorithms.

It should be noted that Simple AR and Matched AR are optimized in trials of [5] to make a compromise between correct and false detection ratios to obtain the highest score. In trials of this paper, algorithms of [6] are tried with different thresholds ($\mu_\alpha = [3 : 0.5 : 10.5], \mu_\beta = [3 : 0.5 : 10.5]$). In most values of thresholds the d' score has converged to the infinity and results for the above mentioned threshold values are taken. The reason of this selection was to be able to interpret the data. Although, the most desirable outcome would be to have $R_{CD} = \%100$ and $R_{FD} = \%0$ which converges to $d' = infinity$ because of the inverse of normal distribution function, it is meaningless to take infinity results containing high false detection ratio.

The high false detection ratio can be interpreted as the algorithms failing to differentiate between percussive sounds and clicks [5]. Although criterion here is not the recovery, when the signals recovered by the algorithms mentioned in [6] and [7] are listened, one can hear that clicks are eliminated with little to no distortion.

In light of these results, one can safely say that the algorithms mentioned in [6] and [7], are too sensitive to solely detect the *perceptible* clicks. However, since these algorithms are shown to be working and successfully detecting artifacts in the signal, these algorithms can be transformed to only detect *perceptual* clicks by a careful selection of the parameter ranges. Simple AR (from [8] and optimized in [5]) has performed the best amongst the compared AR-model based detection algorithm variations.

5. Conclusion

In this paper, 11 different variations of AR-model based click detection algorithms are compared based on their ability to detect only the perceptible clicks. The audio signals used for the trials are categorized, to be containing or not containing clicks, in [5] by listening tests. The comparison criteria was based on the inverse of the normal distribution of the correct and false detection ratios. The results presented show that Simple AR model is best algorithm for the detection of perceptible clicks. Although, all the algorithms are able to detect clicks, the sensitivity of the others resulted in detection of even non-perceptible clicks which in turn created a low or an uninterpretable score.

Acknowledgements

Research described in the paper was supervised by F. Rund, dept. of Radioelectronics, FEE CTU in Prague and supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS20/180/OHK3/3T/13.

References

[1] ESQUEF, P.A., KARJALAINEN, M., VÄLIMÄKI, V., Detection of clicks in audio signals using warped linear prediction, *14th International Conference on Digital Signal Processing Proceedings*, 2002 pp. 1085 – 1088.

[2] DE CARVALHO, H.T., AVILA, F.R., BISCAINHO, L.W.P., Bayesian restoration of audio degraded by low-frequency pulses modeled via gaussian process, *IEEE Journal of Selected Topics in Signal Processing*, 2021, vol. 15, no. 1, pp. 90–103.

[3] LIN, H., GODSILL, S., The multi-channel ar model for real-time audio restoration, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2005 pp. 335 – 338.

[4] NUZMAN, J., Audio restoration: An investigation of digital methods for click removal and hiss reduction, *University of Maryland Institute for Advanced Computer Studies*, 2004.

[5] RUND, F., VENCOVSKÝ, V., SEMANSKÝ, M., An evaluation of click detection algorithms against the results of listening tests, *Journal of the Audio Engineering Society*, 2021, vol. 69, no. 7/8, pp. 586–593.

[6] CIOŁEK, M., NIEDŹWIECKI, M., Detection of impulsive disturbances in archive audio signals, *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017 pp. 671–675.

[7] OUDRE, L., Automatic detection and removal of impulsive noise in audio signals, *Image Processing On Line*, 2015, vol. 5, pp. 267–281, https://doi.org/10.5201/ipol.2015.64.

[8] VASEGHI, S.V., RAYNER, P.J.W., Detection and suppression of impulsive noise in speech communication systems, *IEEE Proceedings I (Communications, Speech and Vision)*, 1990, vol. 137, pp. 38–46(8).

[9] VASEGHI, S.V., RAYNER, P.J.W., A new application of adaptive filters for restoration of archived gramophone recordings, *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*, 1988 pp. 2548–2551 vol.5.

[10] GODSILL, S.J., RAYNER, P.J.W., *Removal of Clicks*, p. 99–134, Springer-Verlag, 1998.

About Authors...

Arda ÖZDOĞRU was born in Turkey. Received his Bachelor's in Electrical and Electronics Engineering from Middle East Technical University in 2021 and currently, attending his Master's of Science in Audiovisual and Signal Processing at Czech Technical University in Prague and conducting research on Signal Processing Related to its Perception.