Low Delay Error Concealment for Audio Signals

Stephan Preihs[1], Fabian-Robert Stöter[1*] and Jörn Ostermann[1]

[1] *Institut für Informationsverarbeitung, Leibniz Universität Hannover, Germany*

Correspondence should be addressed to Stephan Preihs (`preihs@tnt.uni-hannover.de`)

ABSTRACT

This paper is concerned with error concealment in digital audio signals. Two model based methods for signal extrapolation are developed and compared. Among these is the usage of adaptive Kalman filtering which can be processed with zero algorithmic delay in real-time. We also improved an approach which is based on linear prediction by enhancing it with an adaptive parameter usage. The proposed methods, especially the linear prediction approach, perform excellently on maximum gap lengths of about 10 ms. On some monophonic signals the concealment can be extended up to 100 ms without a loss of perceived quality. Both approaches have been compared and evaluated by an informal listening test.

1. INTRODUCTION

Wireless digital transmission systems are vulnerable to burst errors which directly affect the underlying bitstream. If resources for channel coding are limited, the use of error correcting codes is not always feasible. This is especially true for systems which perform under very low delay conditions since channel coding normally would increase their overall latency. If the transmitted signal is an audio signal, errors in the bitstream will be perceived as artifacts. In practice the audio signal therefore is muted for the duration of the gap which only can reduce the artifacts and a perceived degradation of the signal remains.

Error concealment solves this problem by generating a replacement signal that not only fits into the gap but makes it unnoticeable. A real world scenario might require that gaps of up to 100 ms have to be concealed; however, the gap length is of unknown duration at the time of occurrence. In addition, live applications have critical demands for the overall delay of the concealment system that have to be met. Applications could be wireless audio transmission for live applications, real-time jam sessions over the Internet or teleconferencing applications via RTP.

To our knowledge there are no published burst error concealment techniques which were specifically designed to perform under real-time conditions and at the same time do not add any delay. In this paper we assume that the delay constraint is crucial. We focus our research on model based approaches which are able to perform in real-time using only past samples.

2. PROBLEM STATEMENT AND NOTATION

A section of a time discrete digital audio signal can be noted as

$$x(n) = [x_1, x_2, \ldots, x_N]. \quad (1)$$

An error occurring between sampling instants $n = \tau_1$ and $n = \tau_2$ can be described as an assignment of samples x_n to the set of incorrect or missing samples Λ. According to this, the remaining (error free) samples can be assigned to belong to the complementary set $\overline{\Lambda}$ and therefore a signal containing a burst error can be denoted as

$$x(n) = [\underbrace{x_1, \ldots, x_{\tau_1 - 1}}_{\in \overline{\Lambda}}, \underbrace{\hat{x}_{\tau_1}, \ldots, \hat{x}_{\tau_2}}_{\in \Lambda}, \underbrace{x_{\tau_2 + 1}, \ldots, x_N}_{\in \overline{\Lambda}}]. \quad (2)$$

The period $\Delta\tau = \tau_2 - \tau_1 + 1$ between beginning and end of a drop out lies in the range of

$$\Delta\tau_{min} \leq \Delta\tau \leq \Delta\tau_{max}. \quad (3)$$

Following [1] error concealment methods can be divided into two main classes

- Methods that predict and therefore extrapolate the signal only based on past samples
- Methods that interpolate the signal by also using future signal values.

*Work was carried out at Leibniz Universität Hannover. Fabian-Robert Stöter is now with International Audio Laboratories Erlangen.

According to our assumptions the length of the interval $\Delta\tau$ is unknown at the time instant τ_1. The replacement signal therefore can only be extrapolated "from the left". The error concealment is supposed to only cope with burst errors which means that $\Delta\tau_{min}$ is always greater than 1. The upper limit $\Delta\tau_{max}$ was chosen to be about 100 ms. The rate of errors occurring was not limited at first. However, we assume that

$$|\overline{\Lambda}| \gg |\Lambda| \tag{4}$$

applies. The information of a sample $x_n \in \Lambda$ is not used. Therefore we can assume that

$$x_n = \begin{cases} x_n & \text{for } x_n \in \overline{\Lambda} \\ 0 & \text{for } x_n \in \Lambda \end{cases} \tag{5}$$

does not contain arbitrary values. In this work we presume that the information of a sample belonging to one of the sets is present as meta information.

As mentioned before we assume that a minimum delay caused by the error concealment is crucial. In this context delay refers to the total delay between the information about an error being present going into the system and a replacement signal coming out of its output. This delay can be caused by both algorithmic and processing delay. Since we decided to implement an error concealment scheme which can be used as independent building block it does not influence the previous processing and therefore its delay adds up to the total delay of a transmission chain.

Under the assumption that at every time instant it is known if a sample is correct or not we claim that the total algorithmic delay of the concealment system should be 0 samples. This means that it should be possible to compute an alternative value for every sample.

A flow graph of the basic structure of a concealment system can be seen in Figure 1. The functional block "Combination" receives a control signal from the "Error Detection" for switching between the concealment signal and the original signal or doing a fading between them. A "Buffer" of size m allows the error concealment scheme to use statistical dependencies of past samples.

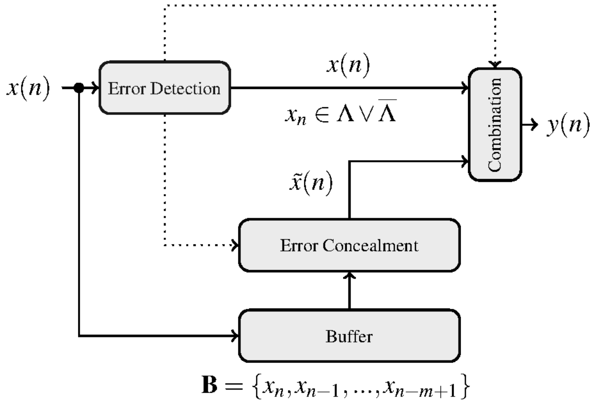

Fig. 1: Basic Principle of a Concealment System.

3. ERROR CONCEALMENT STRATEGIES

Most of the well known methods for error concealment have been developed regarding other requirements than the one considered in this work. In most cases it is assumed that non-overlapping block processing of the signal is possible. Therefore many methods are called *packet loss recovery/concealment*.

In [2] an overview of error concealment methods is given. Using the application as criterion the methods are divided into three main classes - insertion, interpolation and regeneration. We divide the methods into the classes of "trivial", "waveform conserving" and "model based" approaches. Doing this, we do not consider methods that were developed as part of a transmission standard (e.g. ITU G.723.1 or G.729) and therefore are using internal states of the codec.

Trivial approaches do not compute a replacement for corrupted samples but use a predefined signal for error concealment. As shown in [3] replacing corrupted samples by zeros leads to acceptable results for maximum error length $<$ 4 ms and loss rates of less than 2%. Replacing the corrupted signal with white noise even leads to better masking of the error. A repetition of past samples can give good results especially for stationary signals. In packet based transmission systems it is very easy to implement this since it is always possible to replace a missing packet with a previous one.

Waveform conserving methods generate a replacement signal by modification of past sections of the signal. In [4] the basic principle of searching for similar sections in the past signal and doing consecutive copying while trying to avoid discontinuities is presented. Further developments like the ones presented in [5] and [6] also take the actual pitch into account and try to avoid monotony caused by frequent repetition of signal segments. This is done by using multiple overlappings and overlays. In addition [7] further improves this overlap-add by the use of waveform similarity measures. Nevertheless, according

to our baseline studies waveform conserving methods often lead to artifacts that prevent the concealment quality from being perfect.
Modeling of audio signals has proven to be successful for several kinds of audio data compression like redundancy or parametric coding. Therefore it is obvious to also use them for error concealment purposes. One of the most frequently used models for error concealment is the *autoregressive process* (AR-process) applied in a "synthesis by linear prediction" like in [8]. Further developments of this also use pitch information [9], are able to operate in regions of instationarity [10] or to do an extrapolation for longer gaps of stationary signals [11]. Systems that do an AR-based interpolation from both sides of course work better but lead to a delay in the range of the concealment duration. A method called *Smart Copying* [12] is a combination of model based and waveform conserving approaches. It uses a synthetic signal computed on the basis of a prediction residual for the excitation of a synthesis filter.
Several methods do an interpolation of audio signals in other domains like parameter [13] or frequency space [14]. Besides complexity, the problem with these kinds of methods is that the usually used block processing leads to rather high delays.
A quite new approach for signal interpolation is the so called *Audio Inpainting* [15]. It makes use of a description of the signal by a sparse vector and a codebook matrix which can be used for solving an equation system for reconstruction of the original signal. The matrix can have arbitrary entries, is usually overdetermined and often contains basis vectors of the discrete cosine transformation. Using [16] in our preliminary work we found that inpainting is too complex for longer gaps.
Summarizing we can say that none of the mentioned and known methods has proven to be able to cope with our requirements regarding concealment quality, delay and complexity. As a conclusion we decided that for our requirements a model based extrapolation "from the left" is the preferred method.

4. SIGNAL EXTRAPOLATION

According to the results of our baseline studies we developed a model based approach that allows for signal extrapolation only on the basis of past samples and without any delay caused by block processing.
The structure of the rest of this chapter is as follows. In section 4.1 the Kalman filter approach for signal extrapolation is introduced. In 4.2 the underlying signal models are presented, in 4.2.1 two methods for system identification are analyzed. Sections 4.2.2, 4.2.3 and 4.2.4 introduce the conversion, implementation and procedure used for signal extrapolation with the Kalman filter. The influence of stationarity on system identification is discussed in Section 4.2.4. In 4.3 an alternative linear prediction based approach is described which is compared to the Kalman approach in 4.4.

4.1. Kalman Filter Approach

Kalman filtering, which was originally developed for control theory, has been widely used across a variety of engineering applications. In recent years Kalman filtering has also been employed in audio applications. It has been successfully adopted to noise cancellation and speech enhancement problems [17, 18, 19]. In this paper we focus on the use of Kalman filtering for adaptive signal reconstruction and synthesis across signal gaps.
The Kalman filter [20] is an optimal estimator for an unknown system state vector $\mathbf{x_n}$ at the time instant n. The state vector can contain signal values from time domain such as the instationary audio signal but is not limited to consist of amplitude values. The optimality is based on the fact that under certain conditions the state estimation quality of a Kalman filter cannot be surpassed by any other estimator. The output of the Kalman filter is known to be of minimum variance and unbiased. This not only gives great results for noise cancellation but can also be used for a reliable synthesis.
The concept of Kalman filtering is closely related to error concealment due to its nature as a stochastic state observer. In fact the error concealment problem can be viewed as revealing a hidden state for the duration of a gap. At its core it can be separated as two consecutive steps: Based on an underlying system model $\mathbf{A}$ the *prediction* step estimates the current state $\mathbf{x}_n$. The following *update* or *correction* step takes the measurement signal into account. Additionally the Kalman filter setup allows to adjust the reliability of the measured signal. This can be used for morphing from the replacement signal into the original signal instead of a simple linear crossfade at the end of a gap.

4.2. Parametric Signal Model

The system model matrix $\mathbf{A}$ of dimension $(r \times s)$ should model the system state as closely as possible. One of the most often used approaches for modeling audio signals utilizes an AR-process:

$$x(n) = \sum_{k=1}^{p} a_k x(n-k) \tag{6}$$

with $\{a_1, a_2 ..., a_p\}$ being the AR-coefficients. The AR-coefficients can be identified by system identification algorithms like the ones described in the next paragraph. The AR-model has several benefits e.g. it is simple to manage as the physical dimension corresponds to the measured samples in time domain.

A different approach is based on using *sinusoid partials* so that

$$x(n) = \sum_{k=1}^{p} A_k \cdot sin(\omega_k) \tag{7}$$

where ω_k is the frequency of the kth component and A_k its amplitude. This has also been successfully used in low bitrate coding [21, 22].

Both models might become insufficient when the noise level of the input signal is too high. To keep the model order p low both models should be enhanced by a noise model synthesized from the remaining residuum. Details for this will be described in Chapter 5.3.

We did extensive research on adapting the sinusoid model to Kalman extrapolation. Therefore we developed an extension to [23] who presented a multi-tone Kalman frequency tracker. Although we managed to build a working concealment system it performed poorly on real world audio signals because of an instability of this non-linear approach. This led us to dismiss further research on the sinusoid based Kalman extrapolation.

4.2.1. AR System Identification

A model based signal representation requires to adapt its parameters according to the input signal. This determination process is called system identification. It is known that audio signals are only partly weak sense stationary. This makes it even more challenging to adapt the parameters in real time and at the same time maintain stability which is required for usage in a Kalman filter system model.

A common approach for AR-system identification is the *Burg algorithm* which is using the maximum entropy approach and therefore calculates a forward- and backward-prediction error directly from $x(n)$. The poles of the identified process can be characterized with high spectral resolution. Even with challenging highly instationary signals and at low filter orders the poles of the identified system describe a stable synthesis filter. Unfortunately the Burg algorithm is based on block processing which requires to adapt the initialization of the Kalman filter to match the low delay requirement.

To avoid block processing adaptive filtering algorithms can be used. Basically most methods belong to the LEAST MEAN SQUARE (LMS) or RECURSIVE LEAST SQUARE (RLS) class of algorithms. Although LMS identification approaches are computationally efficient and do have excellent tracking capabilities, they turned out to be unstable for some signals especially if used with high filter orders.

Since they turned out to have great stability behavior for the identified poles we chose the RLS type approaches for system identification. The forgetting factor $\lambda = \frac{1}{1-\tau}$ allows to adjust the memory of the algorithm to match a finite number of samples τ. Although the standard RLS algorithm is computationally expensive there do exist faster implementations like FAST-RLS which could be implemented with complexity of $\mathcal{O}(n)$ [24].

We compared block based identification (BURG) with an adaptive filter based one (RLS) and found out that despite their fundamental differences they both can be used in Kalman filter system models. The Chapter Extrapolation will illustrate both cases and their customized concealment scheme.

4.2.2. State-Space Representation

One of the key characteristics for Kalman filtering is the dependency on state-space representation for all input and output signals. The state vector $\mathbf{x}$ does have a dimension of size $(r \times 1)$ and represents the internal state of a dynamic system. In addition to the state vector the output vector $\mathbf{y}$ which holds the measurement or observation has dimension $(s \times 1)$. The state-space representation now allows for separating the system state from the observation, hence equations 8 represent the common state-space observer which is described by

$$\begin{aligned} \mathbf{x}_n &= \mathbf{A}\mathbf{x}_{n-1} \\ \mathbf{y}_n &= \mathbf{H}\mathbf{x}_n \end{aligned} \tag{8}$$

with matrix $\mathbf{A}$ of dimension $(r \times r)$. $\mathbf{H}$ is of dimension $(s \times r)$ and connects the measurements with the one-step ahead prediction $\mathbf{x}_n$ of the system state. The output $\mathbf{y}_n$ can then be set to the amplitude values of the audio signal.

The AR-coefficients a_k found by the presented system identification can easily be inserted into the system matrix $\mathbf{A}$. Although there are multiple ways for representation we have chosen the common frobenius form. With

$r = s = p$ the resulting system matrix $\mathbf{A}$ is:

$$\mathbf{A} = \begin{bmatrix} a_2 & \cdots & a_{p-2} & a_{p-1} & a_p \\ 1 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

4.2.3. Kalman filtering

The actual Kalman filtering has been implemented using separate prediction and update steps. The equations and procedure will not be detailed here as we used standard implementations in Matlab [25]. The system model order p is fixed. The system modeling error $\mathbf{Q}$ and the measurement error $\mathbf{R}$ of the Kalman filter need to be adapted to the signal modeling. Matrix $\mathbf{Q}$ can be interpreted as the difference between the actual dynamic system and its system model representation. Matrix $\mathbf{R}$ can be used to control how far the measurements can be trusted.

For the AR-model, $\mathbf{Q}$ can be taken from the system identification. E.g. in case of the Burg method we get it directly by calculating the variance of the prediction error $e(n)$ for the reflection coefficients of the past p samples so that $\mathbf{Q}(\mathbf{n}) = \mathbf{I} \cdot Var(e(n))$.

In case when no error occurs we can measure the amplitude directly from x(n) in every step. To reflect this relation, the measurement matrix is set to

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$$

so that only the most recent samples are used for updating. When using amplitude values within the system state it seems that the measurement noise is non-existent. In a real implementation $\mathbf{R}$ should be set to a very small value, so that the Kalman filter does not make the prediction depend directly on the measurement and the system dynamics are kept high.

For the AR-model the order p of the Kalman filter should be matched to the order of the AR-process. Both, BURG and RLS algorithm share the fact that the prediction error will be smaller for larger p on complex audio signals. We will determine a matched order in Chapter 5.1.

4.2.4. Concealment Scheme and Quasi-Stationarity

Using Kalman filtering to provide a replacement signal for every sampling instant requires a custom concealment scheme. Fortunately this can easily be accomplished by adjusting the measurement error covariance $\mathbf{R}$: If at sampling instant $n = \tau_1$ an error occurs $\mathbf{R}$ should be raised to a finite maximum value. The result is, that the measured input value will hardly be used within the update step of the Kalman filter. To avoid numerical instabilities it can be helpful to even completely skip the update step during the gap. The Kalman filter then is set into prediction-only mode. When the signal at time instant $n = \tau_2$ is back $\mathbf{R}$ could be set back to its original close-to-zero value. This approach therefore also enables to morph between original signal and prediction by fading $\mathbf{R}$. The Kalman filtering concealment procedure with skipping of the update step during the gap is outlined in Figure 2.

It can be shown that a system modeled by an AR-process is only observable when the process to be observed is at least weak sense stationary [26]. The optimality condition of the Kalman filter only persist if both error covariances are known and if the error is normal distributed and zero mean. This might not be the case for typical audio signals. We therefore considered two options for adjusting the concealment scheme as follows.

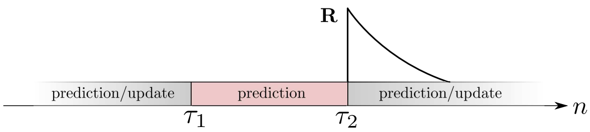

Fig. 2: Kalman Filtering Concealment Procedure with Covariance Alteration.

Filtering Instationary Signal (RLS Identification) We found out that the Kalman filter can be set up to process real (instationary) audio signals without diverging. This requires to change the system model for every sample instant, so that the matrices become $\mathbf{A}(\mathbf{x}, n)$ and $\mathbf{H}(\mathbf{x}, n)$. The RLS adaptive identification approach provides an updated model every sample instant. The RLS forgetting factor λ allows for the AR-coefficients to vary slowly which helps the Kalman filter to retain stable and responsive. This turns the system into a non-linear approach as it is now a combined state and parameter estimator.

Using this dual filter approach we are able to achieve stable extrapolation for real audio signals.[1] In fact a similar approach has been taken by [28] to denoise audio signals. This combination avoids the use of the Extended Kalman Filter. We adjusted the error covariances experimentally for avoiding instabilities within the output signal which

[1] For the relationship between RLS and Kalman filtering see [27].

could be caused by instationarities. The treatment of remaining audible artifacts (e.g. for speech signals during longer gaps) will be addressed in Chapter 5.3

Filtering Stationary Signal (Burg Identification) To minimize the problems of fast changing signals, the input has to be at least partly stationary. If an error occurs at time instance τ_1 our proposed concealment scheme therefore is as follows:

- System identification for order p out of previous stationary n_s samples (stationarity determination comparable to 5.1)
- Start Kalman filtering with the identified model beginning with time instant $n = \tau_1 - n_s + 1$
- Filtering across the gap.

This requires additional computational cost but allows for stable system identification using the common Burg algorithm.

4.3. Linear Prediction Approach

The Kalman filter approach which utilizes parts of stationary signals is quite similar to linear prediction based error concealment using IIR filter synthesis. [29] proposed a similar concealment scheme based on a simple IIR Filter excited with zeros. In fact our Kalman filter approach using Burg algorithm gives nearly the same results as the linear prediction based approach. The slight differences come from the fact that the Kalman filter approach is stochastic and the linear prediction deterministic. We benchmarked and compared both results using a modified Monte Carlo run and explored that the differences were inaudible by human ear.

4.4. Discussion

Our proposed Kalman filter concealment approach is very flexible as it is based on signal models. We think that it has some advantages over standard linear prediction based approaches e.g. [29].

- Kalman filtering allows for a fully adaptive concealment system.
- It can take advantage of every correctly transmitted sample so that the system can perform even on very short burst errors or high bit error rates.
- Adaptive recursive systems are well suited for real-time and very low delay applications. In fact our approach works without any additional delay.
- Our approach can extrapolate gaps of several hundred milliseconds without any external excitation signal.
- Using the proposed RLS method it can be shown that non-stationary signals do not raise the complexity of the concealment method.
- This approach can even be used for extrapolation of very noisy signals since the Kalman filter can be used for de-noising without adding any additional computational cost.

We also want to note that the extrapolation quality of the RLS based approach is worse than using the Burg approach due to many safety measures. Evaluation of the perceived quality will be presented in Chapter 6.

5. INTEGRATION IN THE OVERALL CONCEALMENT SYSTEM

Figure 3 shows a flow graph of the over-all concealment system. The system can be divided into four main building blocks named Synthesis Filter (either Kalman filter or Linear Prediction Approach), System Order Determination (only Linear Prediction Approach), Break-off Condition and Noise Modeling with the latter three being described in the following.

5.1. Stationarity Determination and Variable System Order (Linear Prediction Approach)

As described in 4.3 one possible approach for signal synthesis is the Linear Prediction Approach. Using a block based pre-processing allows for an optimal adaption of the concealment to the signal. This can be further improved by the usage of a variable system order and adaptable input blocksize for the system identification. We choose the blocksize of past samples used for system identification according to the stationarity of the signal. The number of signal samples that can be accounted as quasi stationary are calculated by the spectral distance which leads to a stationarity index vector $STI(n)$ (see eq. 9). We determine the kolmogorov distance [30] by windowed Fourier transforms calculated and shifted section wise. The resulting time series is low pass filtered and a search for peaks is done. The number of samples stored for the last maximum is used for system identification.

$$STI(n) = \|X(m,\omega)\|_{kolmogorov} \tag{9}$$

A matched variable system order is used to take into account the number of spectral components of the signal.

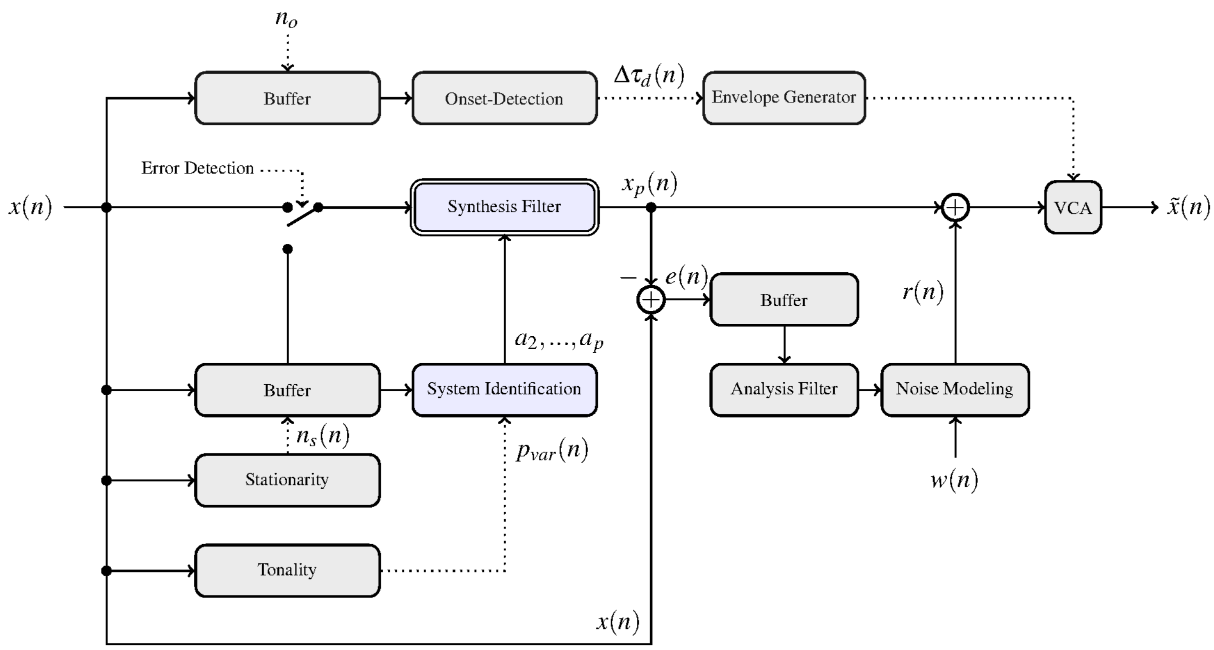

Fig. 3: Flow Graph of the whole Concealment System.

Unlike [31] we found that for tonal signals a system order in the range of 200 is sufficient but for vocals and speech higher values are needed. Therefore we experimentally found the relationship

$$p_{var} = max((1-t)\cdot 1000, 250) \qquad (10)$$

for the system order where t describes the tonality of the signal excerpt and is in the range of 0 to 1.

5.2. Break-Off Criterion

As mentioned in the introduction we aim at concealment of maximum gap length $\Delta\tau_{max}$ in the range of 100 ms. However, for some signal classes partial muting helps improving the perceived quality more than a continuation of the signal. Therefore we introduced a decay variable $\Delta\tau_d < \Delta\tau_{max}$ that controls the decline of signal energy in the concealment case. For computation of $\Delta\tau_d$ a mean signal activity period is estimated from n_o past samples by detecting pauses in between signal parts. This is done by searching for minima in the spectral flatness. With $\Delta\tau_{n,min}$ as the number of samples since the last local minimum and $E[n_{active}]$ as mean signal activity period the remaining concealment time is determined by

$$\Delta\tau_{remain}(n) = E[n_{active}] - \Delta\tau_{n,min}. \qquad (11)$$

For further reducing the concealment artifacts of speech signals we also take the spectral spread $S_{spread} = Var(|H(j\omega)|)$ into account. The final concealment duration therefore is computed by

$$\Delta\tau_d(n) = min(\Delta\tau_{remain}(n) - a\cdot S_{spread}, \Delta\tau_{max}). \qquad (12)$$

with a in the range between 3000 and 5000.

5.3. Noise Modeling

Besides frequency and temporal properties one of the features that should be reflected by an error concealment is the amount of noise present in a signal. For a noisy tonal signal a lack of noise within the concealed gap might prevent from a perfect concealment. Therefore we model the noise components which significantly influence the auditory impression. This is done on the basis of a residual signal $x(n) - x_p(n)$ which is taken from the concealment system, stored in a buffer, sectionally filtered by a bark filter bank and analyzed afterwards. Assuming that the spectral distribution of residuum is only changing slowly, the mean power of the filtered residual signal can be used for computation of an equivalent non-artificial noise source. The synthesis is then realized by weighted addition of filtered non-artificial noise $w(n)$ stored in lookup tables which are played back in alternating direction for avoiding similar noise patterns in consecutive concealment cases.

6. EVALUATION

For evaluation of subjective concealment quality an informal listening test with 11 non-expert listeners was conducted in a quiet demonstration room. Although being designed for audio codec evaluation we chose the MUSHRA methodology [32] as basis for the test design. The graphical user interface was implemented on a personal computer and worked like the reference implementation in the standard. A high quality digital to analog converter and STAX headphones/amplifier were used. The test items were based on the MPEG and SQAM database as well as self provided recordings of critical sounds. They were chosen to be from the three signal classes instruments, vocals and speech and are of CD quality sampling rate/bit resolution.

Item	Source
Trumpet	EBU Test set
Glockenspiel	MPEG Test set
Acoustic guitar	Self provided
Sopran, female	EBU Test set
Vocals, pop, male	Self provided
Speech, male, German	MPEG Test set

Table 1: Items for the listening test

The tested conditions are the Kalman Filter Approach (KF) using RLS for system identification as described in 4.2.4, the Linear Prediction approach (LPC) with variable system order according to 4.3 as well as the hidden reference and an anchor which is the signal just containing gaps without fading at the beginning or end of a gap. For testing and optimization of the error concealment approaches different (partly randomized) error models were used since this is the only way for realistic performance evaluation. However, for the listening test we decided to use a constant spacing of 1 s between error positions. This was done for making the comparison easier for the subjects and also for avoiding that some of the errors are not perceived by them. We used two different error length of 10 ms and 100 ms which leads to 12 signals.

Figure 4 shows the results of the listening test for each item and the different tested conditions as mean values and 95% confidence intervals in the order of presentation[2] to the subjects.

The results show that both concealment strategies are almost always better than muting the signal during a burst error. The Linear Prediction Approach gets very good scores for some signals even for 100 ms. For 10 ms gap length some of the subjects also mixed-up the reference and concealed signals. The Kalman Approach allows for about 50 % of improvement for most signals but overall has a worse performance than the Linear Prediction Approach. The reason for this are artifacts in the concealment that arise during extrapolation of instationary signal parts. The rather big confidence intervals show that these artifacts are judged differently by the subjects. Nevertheless for some signals like the Glockenspiel the Kalman Approach still achieves good results. As can be seen from the results of both gap lengths, speech is always critical regarding concealment quality. In addition the subjective impression differs between subjects which again is a reason for the rather large confidence intervals.

7. CONCLUSION AND OUTLOOK

In this paper two model based approaches for no-delay error concealment in digital audio signals by signal extrapolation are presented and compared. The first approach is based on adaptive Kalman filtering which can be processed in real-time and sample based without any block processing. The second approach uses the well known Linear Prediction Approach and is extended to include a variable system order. Both approaches are integrated into a concealment system that incorporates a break-off criterion for partial muting and a noise modeling for concealment of noisy signals.

The different versions of the system have been evaluated by an informal listening test. The results show that both proposed methods perform very good on gap lengths of about 10 ms. For some monophonic signals, especially for the Linear Prediction Approach, the concealment can be extended up to 100 ms without a loss of perceived quality. In addition the Linear Prediction Approach allows for a near transparent concealment for tonal signals. The Kalman Filter approach is not always able to compete against the performance of the Linear Prediction Approach because of artifacts occurring for instationary signals. Nevertheless it still helps to improve the

[2] Contrary to the recommendation in the standard we decided not to randomize the order of presentation of the different items for being able to see learning effects (e.g. one can see that some of the subjects did mix up the reference with the Linear Prediction Approach for the first item).

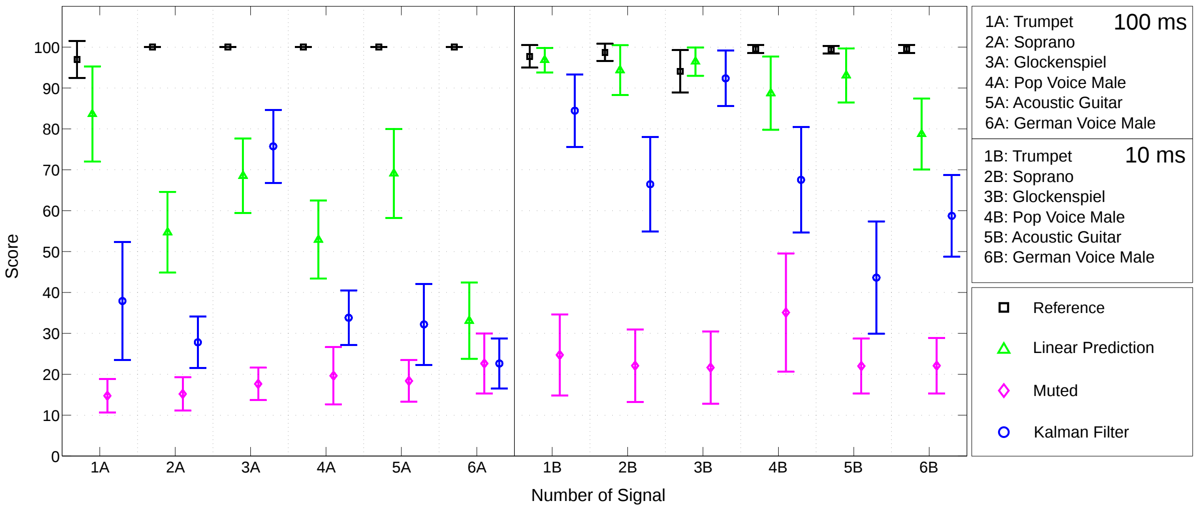

Fig. 4: Results of the MUSHRA listening test with 11 listeners (mean values and 95 % confidence intervals).

subjective quality of a corrupted signal significantly.
With the Kalman Filter Approach we were therefore able to show that real-time sample based error concealment without delay is possible, though there are remaining quality improvements to be made. One possible approach could be to combine multiple signal models by an *Interacting Multiple Models Kalman filter* [33] so that it can be adapted to different signal classes. As we showed in our work an optimized break-off criterion can result in better quality performance. Especially for concealment of speech and instrumental signals the parameters could be adapted to better fit this dual use-case.

8. REFERENCES

[1] S. Vaseghi, *Multimedia signal processing: theory and applications in speech, music and communications*. J. Wiley, 2007.

[2] C. Perkins, O. Hodson, and V. Hardman, "A survey of packet loss recovery techniques for streaming audio," *IEEE Network*, vol. 12, no. 5, 1998.

[3] N. Jayant, "Effects of packet loss on waveform coded speech," in *Proc. Fifth Int. Conference on Computer Communications*, 1980.

[4] D. Goodman, G. Lockhart, O. Wasem, and W. Wong, "Waveform substitution techniques for recovering missing speech segments in packet voice communications," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 6, 1986.

[5] O. Wasem, D. Goodman, C. Dvorak, and H. Page, "The effect of waveform substitution on the quality of pcm packet communications," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 3, 1988.

[6] D. Griffin and J. Lim, "Signal estimation from modified short-time fourier transform," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, no. 2, 1984.

[7] W. Verhelst and M. Roelands, "An overlap-add technique based on waveform similarity (wsola) for high quality time-scale modification of speech," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, 1993.

[8] A. Janssen, R. Veldhuis, and L. Vries, "Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 2, 1986.

[9] S. Vaseghi and P. Rayner, "Detection and suppression of impulsive noise in speech communication systems," in *Communications, Speech and Vision, IEE Proceedings I*, vol. 137, no. 1. IET, 1990.

[10] W. Etter, “Restoration of a discrete-time signal segment by interpolation based on the left-sided and right-sided autoregressive parameters,” *Signal Processing, IEEE Transactions on*, vol. 44, no. 5, 1996.

[11] I. Kauppinen and J. Kauppinen, “Reconstruction method for missing or damaged long portions in audio signal,” *J. Audio Eng. Soc*, vol. 50, no. 7/8, 2002.

[12] M. Niediwiecki and K. Cisowski, “Smart copying-a new approach to reconstruction of audio signals,” *IEEE Transactions on Signal Processing*, vol. 49, no. 10, 2001.

[13] R. Maher, “A method for extrapolation of missing digital audio data,” *Journal of the Audio Engineering Society*, vol. 42, no. 5, 1994.

[14] M. Lagrange, S. Marchand, and J.-B. Rault, “Long interpolation of audio signals using linear prediction in sinusoidal modeling,” *Journal of the Audio Engineering Society*, vol. 53, no. 10, 2005.

[15] A. Adler, V. Emiya, M. Jafari, G., M. Elad, R. Gribonval, and M. D. Plumbley, “Audio Inpainting,” INRIA, Tech. Rep. RR-7571, Mar. 2011.

[16] V. Emiya, A. Adler, and M. Jafari, “Audio inpainting toolbox,” http://small-project.eu/software-data/AudioInpaintingToolbox.zip/view.

[17] S. Gannot, D. Burshtein, and E. Weinstein, “Iterative and sequential kalman filter-based speech enhancement algorithms,” *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 4, 1998.

[18] M. Gabrea, “Robust adaptive kalman filter for speech signal recovery in colored noise,” in *International Workshop on Acoustic Echo and Noise Control (IWAENC 2005)*, 2005.

[19] G. Doblinger, “An adaptive kalman filter for the enhancement of noisy ar signals,” in *IEEE International Symposium on Circuits and Systems, ISCAS ’98.*, vol. 5, may-3 jun 1998.

[20] R. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of basic Engineering*, vol. 82, no. Series D, 1960.

[21] R. McAulay and T. Quatieri, “Speech analysis/synthesis based on a sinusoidal representation,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 4, 1986.

[22] B. Edler and H. Purnhagen, “Parametric audio coding,” in *5th International Conference on Signal Processing Proceedings, 2000*, vol. 1. IEEE, 2000.

[23] D. Alves and R. Coelho, “An adaptive algorithm for real-time multi-tone estimation and frequency tracking of non-stationary signals,” *IEEE Transactions on Nuclear Science*, no. 99, 2011.

[24] J. Proakis, C. Rader, F. Ling, C. Nikias, M. Moonen, and I. Proudler, *Algorithms for statistical signal processing*. Prentice Hall, 2002.

[25] S. Särkkä, J. Hartikainen, and A. Solin, “Ekf/ukf toolbox,” Helsinki University of Technology. [Online]. Available: http://www.lce.hut.fi/research/mm/ekfukf

[26] P. Brockwell and R. Davis, *Time series: theory and methods*. Springer Verlag, 2009.

[27] S. HAYKIN, *Adaptive Filter Theory*, third edition ed. Prentice Hall, 2001.

[28] D. Labarre, E. Grivel, M. Najim, and E. Todini, “Two-kalman filters approach for unbiased ar parameter estimation from noisy observations, application to speech enhancement,” in *EURASIP-EUSIPCO*, Wien, 2004.

[29] I. Kauppinen and K. Roth, “Audio signal extrapolation - theory and applications,” *Proc. of the 5th Int. Conference on Digital Audio Effects (DAFx-02)*, 2002.

[30] H. Laurent and C. Doncarli, “Stationarity index for abrupt changes detection in the time-frequency plane,” *Signal Processing Letters, IEEE*, vol. 5, no. 2, 1998.

[31] I. Kauppinen and K. Roth, “An adaptive technique for modeling audio signals,” in *Conference on Digital Audio Effects DAFX*, 2001.

[32] *Method for the subjective assessment of intermediate quality level of coding systems BS.1534-1*, ITU-R, 2003.

[33] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, “Interacting multiple model methods in target tracking: a survey,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 1, 1998.