ADPCM SPEECH CODER WITH ADAPTIVE TRANSMISSION AND ARMA MODELLING OF NON-UNIFORMLY SAMPLED SIGNALS

Élisabeth Lahalle and Jacques Oksman

ÉCOLE SUPÉRIEURE D’ÉLECTRICITÉ, Service des Mesures
Plateau de Moulon, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette Cedex, FRANCE
Elisabeth.Lahalle@Supelec.fr, Jacques.Oksman@Supelec.fr

ABSTRACT
In a previous paper [1], a coding principle based upon non-uniform transmission of signal samples was introduced. This coding principle relies on the use of an efficient prediction/reconstruction algorithm for non uniformly sampled signals in both the transmitter and the receiver. Recently, we proposed an adaptive ARMA estimation method for time series with missing samples [2]. The purpose of this paper is to study the feasibility of an ADPCM coding scheme using adaptive transmission in order to reduce the average bit rate, that could allow to share multiple lines carrying voice or to multiplex voice and data. The ITU-T G.726 ADPCM standard [3] is considered in order to evaluate the performances of the proposed speech coder.

1. INTRODUCTION
In many physical systems, signals are sampled in a deliberately non-uniform manner. Such cases are found, for example, for power saving or, more generally, for data compression purpose. The key idea of the data compression technique proposed in [1] is to transmit the samples "only when necessary" combining prediction and quantization. This relies on the following: An efficient prediction may provide a smaller error than quantization at numerous sampling instants. Therefore, reduced average transmission rate, for a similar signal to noise ratio, may be achieved by "efficient" prediction of signal samples during some "convenient" sampling instants. The aforementioned principle is applied to the design of an ADPCM speech coder able to transmit simultaneously different signals.

In section 2, after having briefly described the previous methods on which the present work is based, and the G.726 ADPCM coding standard, we present a new coding scheme. Simulation results show the validity of the proposed coding scheme in section 3.

2. PROPOSED METHOD

2.1. Context
2.1.1. Compression principle
In [1], it was shown that an efficient prediction method for non-uniformly sampled signals could allow to replace a transmitted quantized sample by a predicted one at numerous sampling instants, giving a smaller or similar reconstruction error at the decoder output. Applying these key ideas to ADPCM coders may lead to the design of coders with adaptive transmission.

2.1.2. ARMA adaptive predictor for Non-Uniformly Sampled Signal (NUSS)
For an ARMA($N_a$, $N_b$) adaptive predictor in the case of non-uniformly sampled signals, the predicted signal $\hat{s}_{n+1,L}$, built with the $L$ last samples, may be written as follows [2]:

$$\hat{s}_{n+1,L} = H_{n+1,L} \theta,$$

with:

$$H_{n+1,L} = \begin{bmatrix}
  h_n & \ldots & h_{n-L} \\
  \vdots & \ddots & \vdots \\
  h_{n-N_a+1} & \ldots & h_{n-L-N_a+1} \\
  h_n - \theta^\top h_n & \ldots & h_{n-L} - \theta^\top h_{n-L} \\
  \vdots & \ddots & \vdots \\
  h_{n-N_b+1} & \ldots & h_{n-L-N_b+1} \\
  -\theta^\top h_{n-N_b+1} & \ldots & -\theta^\top h_{n-L-N_b+1}
\end{bmatrix}.$$

Since the most straightforward way is to replace the missing value by its estimate, we have:

$$h_{n-j} = \begin{cases}
  s_{n-j} & \text{if sample } n-j \text{ is known,} \\
  \hat{s}_{n-j} & \text{if sample } n-j \text{ is lost.}
\end{cases}$$
and:
\[
\begin{bmatrix}
h_{n-1} \\
h_{n-N_a} \\
h_{n-1} - \theta^T h_{n-1} \\
\vdots \\
h_{n-N_b} - \theta^T h_{n-N_b}
\end{bmatrix} = \begin{bmatrix} h_{a,n} \\ h_{b,n} \end{bmatrix}
\]

The ARMA NUSS identification algorithm is summarized as follows:
- if the sample is available, update parameters according to a gradient optimization formula [5] [2](2) (3),
\[
\begin{aligned}
\theta_{a,n+1} &= \theta_{a,n} - \mu_a \frac{\partial J_{n+1,L}}{\partial \theta_a} \\
\theta_{b,n+1} &= \theta_{b,n} - \mu_b \frac{\partial J_{n+1,L}}{\partial \theta_b}
\end{aligned}
\]
with:
\[
J_{n+1,L} = \frac{1}{L} (s_{n+1,L} - \hat{s}_{n+1,L})^T (s_{n+1,L} - \hat{s}_{n+1,L})
\]
- otherwise, replace the missing sample by its estimate.

2.2. IUT-T G.726 ADPCM

The quantizer is the one used in the G.726 at the 32 kbit/s standard. The criterion used to decide of the transmission of the 4 bit code word \( I[k] \) and the 1 bit transmission flag \( T[k] \) (i.e. the extra bit needed to make the difference between transmitted and not transmitted samples in a continuous bit stream) is:
\[
| d[k] | \geq c2^{-y[k]}
\]

2.3. NUSS ADPCM

The proposed ADPCM principle scheme is presented in Figure 3 and Figure 4. This scheme implements the above principle: A sample is transmitted or not according to the value of the prediction error (computed in both the receiver and the transmitter).

The quantizer is the one used in the G.726 at the 32 kbit/s standard. The criterion used to decide of the transmission of the 4 bit code word \( I[k] \) and the 1 bit transmission flag \( T[k] \) (i.e. the extra bit needed to make the difference between transmitted and not transmitted samples in a continuous bit stream) is:
\[
| d[k] | \geq c2^{-y[k]}
\]
threshold which is adapted to the variance of the prediction error for a negligible increase of complexity as compared with the G.726 standard.

3. MAIN RESULTS

3.1. Test signals

In order to evaluate the performances of the coder, both English and French sentences are used in the tests. The results below come from "Mary as a little Lamb" and "Un chasseur sachant chasser doit savoir chasser sans son chien".

3.2. Parameter settings

For the NUSS predictor, all simulations were carried out with adaptation steps $\mu$ satisfying the classical stability conditions [4] [5]:

$$0 < \mu_a N_a \sigma_s^2 + \mu_b N_b \sigma_d^2 < 1$$  \hspace{1cm} (5)

The same $(N_a, N_b)$ predictor order has been chosen for both NUSS and G.726 predictors. Its value is $(2, 6)$ as specified in the G.726 recommendation.

As mentioned before, different bit rates may be achieved for the proposed coding scheme using the value $c$. This scale factor has been chosen in a range allowing to get a mean bit rate between 40 kbit/s (all samples transmitted case) and a mean bit rate more than twice lower than for the G.726 at 32 kbit/s. The results are interesting down to 16 kbit/s as it may be seen in Figures 5 and 6.

3.3. Comparisons

Firstly, an improvement in both bit rate and speech quality (refering to signal to noise ratio) as compared to the G.726 at 32 kbit/s standard may be noticed, in Figure 5, up to a mean bit rate of 23 kbit/s. This can be explained by the use of a better predictor than the sign-sign algorithm [5] [7] [8] introduced in the G.726 standard in order to decrease the complexity of the implementation. More interestingly, it may also be seen, Figure 5 and Table 1, that, for a similar speech quality (1 dB loss for the SNR compared to the G.726 standard at 32 kbit/s), the mean bit rate may be reduced by a factor 2. That may be used to transmit two signals on the same channel. This is not allowed by the G.726 at 16 kbit/s due to its very low SNR (Table 1).

**Table 1: SNR (dB) for "Mary has a little lamb".**

<table>
<thead>
<tr>
<th>Coder</th>
<th>Mean bit rate (kbit/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUSS ADPCM</td>
<td>14</td>
</tr>
<tr>
<td>G.726 ADPCM</td>
<td>13.2</td>
</tr>
</tbody>
</table>

Although the performances are a little below as those of the G.726 at 32 kbit/s in the case of the French sentence, Figure 6 and Table 2, for 1.5 dB loss (as compared to the G.726 standard at 32 kbit/s) the mean bit rate may also be reduced by a factor 2.

**Figure 5.** SNR of the proposed coder versus mean bit rate for ”Mary has a little lamb”.

**Figure 6.** SNR of the proposed coder versus mean bit rate for ”Un chasseur sachant chasser doit savoir chasser sans son chien”.
Mean bit rate (kbit/s) | 32 | 16
---|---|---
Coder | NUSS ADPCM | 14.5 | 13.9
      | G.726 ADPCM | 15.5 | 1.8

Table 2: SNR (dB) for "Un chasseur sachant chasser doit savoir chasser sans son chien".

Figure 7 shows an example of the quality of the waveform reconstruction for the beginning of the "M" of "Mary has a little lamb" transmitted with a mean bit rate of 16 kbit/s. The observation of the transmission flag shows that the transmission rate depends on the local variance of the signal (via the prediction error variance). So, due to the performance of the predictor, the transmission rate may be greatly reduced for non voiced sounds (with a small local variance).

Figures 8 and 9 show another example of the quality of the waveform reconstruction with an important reduction of the transmission rate for a French nasal sound (i.e. the end of the "un" in the French phrase). At many sampling instants, a smaller reconstruction error may be noticed for the NUSS ADPCM than for the G726 ADPCM, even when no sample is transmitted.

4. CONCLUSION

The proposed ADPCM coder allows to transmit simultaneously two signals on the same channel. That is obtained with a small decrease of the speech quality and a small increase of complexity compared to the G.726 standard. The greatest part of the complexity is due to the equations giving the gradient of the NUSS predictor, namely \(2N^3 + 2N^2\) for each missing sample and \(3N^3 + 4N^2 + N\) for each available sample (with \(N = N_a + N_b\)) [2]. The introduction of simplified NUSS adaptive predictors should be an answer to the complexity of the NUSS ADPCM [9].
References


