SOURCE CONTROLLED CHANNEL DECODING FOR GSM-AMR SPEECH TRANSMISSION WITH VOICE ACTIVITY DETECTION (VAD)

C. Murali Mohan  
R. Aravind  

Department of Electrical Engineering  
Indian Institute of Technology, Madras  
ee02s005@iitm.ac.in, aravind@tenet.res.in

ABSTRACT

Ideally, in general any speech encoder should completely remove the redundancy which is present at the input, still there is some residual redundancy remaining in coded speech data even after applying a powerful speech codec. The main goal of Source Controlled Channel Decoding (SCCD) is to improve the performance of the channel decoder by using this redundancy of speech encoded data. Till now SCCD has been applied to GSM Full Rate and Enhanced Full Rate by exploiting the inter and intra frame correlations. In this paper, we describe the use of SCCD for all eight modes of GSM-AMR system. Speech codec of GSM-AMR has the advantage of controlling Discontinuous Transmission (DTX), using the Voice Activity Detector (VAD). We observed interframe correlation in encoded speech and updated silence frames independently. A channel decoder which exploits these interframe correlations and performs better than maximum likelihood sequence estimation (VA) is a priori Soft Output Viterbi Algorithm (APRI-SOVA). Hence we use APRI-SOVA as a AMR channel decoder which exploits the residual redundancy present in the speech coded data.

1. INTRODUCTION

Considerable attention has been drawn to the digital transmission techniques over mobile radio channel, because of the growing importance to the wireless communication systems, but the overall performance is reduced because of noise, interference, multi-path and fading, which are very common with mobile radio channel. To combat this problem convolutional coding is used for the output of speech encoder.

A method for decoding convolutional codes that outperforms the Viterbi algorithm, called Source Controlled Channel Decoding (SCCD), was proposed by Hagenauer [1]. SCCD uses the residual redundancy present at the output of speech encoder. Ideally speech encoder should completely remove the correlation present in the input speech data, but no practical speech coder achieves this. Due to practical constrains like complexity, delay and non-stationarity of speech, some correlation [5] present at the output of speech encoder. The underlying idea with the SCCD consist of observing that between two consecutive frames, the most significant bits do not change very often. This residual inter frame correlation can be used at the receiver by means of an a priori Soft Output Viterbi Algorithm (APRI-SOVA) decoder.

Hagenauer [1] applied SCCD to the GSM Full Rate system using APRI-SOVA by making use of source residual redundancy and showed the performance improvement in terms of BER. Hindelang [2] proposed a slight modification to the HUK model presented by Hagenauer [1] and implemented this modified HUK model to the GSM Full Rate system. Russcitto [3] in his paper described a channel decoding algorithm suitable for the exploitation of intra-frame correlation, by using an approach similar to SCCD.

Another speech encoder which exhibits significant interframe correlation, and hence a potential candidate on which SCCD can be applied is GSM-AMR [5]. The speech codec of AMR is capable of handling eight different modes. Each mode corresponds to a specific bit rate. In this paper we focused on GSM-AMR speech transmission with Voice Activity Detection (VAD) using SCCD. For implementing SCCD we used APRI-SOVA along with modified HUK model.
II. GSM-AMR SPEECH AND CHANNEL CODING

The GSM-AMR speech and channel coding schemes are briefly explained here. The AMR codec can be operated at eight bit rates - 12.2, 10.2, 7.95, 7.4, 6.7, 5.9, 5.15 and 4.75 kbps. The codec is based on CELP coding model [5]. The coder operates on speech frames of 20 ms corresponding to 160 samples at the sampling frequency of 8000 samples/sec.

Generally speech transmission contains considerable amount of silence. To make use of this fact GSM-AMR has a special mode called VAD, with which silence frames can be differentiated from speech frames and coded more efficiently. It does the source coding [5] differently for speech and silence frames, so that the average output bit rate is reduced. For every 8 silence frames the codec produces one frame with comfort noise (CN) parameters, remaining 7 frames contain no data.

After speech coding each 20 ms speech frame contains different number of bits for different modes. According to the subjective importance bits of each speech frame are divided into two classes [6]. The most important bits are put into class 1a which are very sensitive to errors, remaining into class 1b. Each silence frame contains 35 bits only.

For error protection, the GSM-AMR channel coding [6] is done as shown in Figure 1. In all modes class 1a bits of speech frame are protected with CRC of 6 bits, which will be used for bad frame detection at the receiver side. Together with the class 1b bits they are reordered and encoded by a convolutional code with memory l and rate 1/n. See Table 1 for l and n values of all modes. At the end of the channel coding few bits are punctured to make the output bit rate constant (22.8 kbps) irrespective of the mode.

In silence frame all 35 CN bits are important bits. They are protected with a CRC of 14 bits. Then they are encoded by a convolutional code with memory 4 and rate 1/4. Together with identification marker and inband data they are reordered. Finally each silence frame after channel coding contains 456 bits (22.8 kbps).

As mentioned above, the output of the speech coder contains residual redundancy. One such redundancy is interframe correlation which is the correlation between the bits with the same index position q in successive frames. Figure 2 shows two successive frames where the qth bit of kth frame is denoted by $u_{k,q}$. This interframe correlation of each encoded data bit can be measured, both for speech and silence frames by estimating the probability of change, $P_c$. If $u \in \{-1, +1\}$, we consider the change as ($u_{k-1,q} = -1$ and $u_{k,q} = +1$) or ($u_{k-1,q} = +1$ and $u_{k,q} = -1$). Figure 3 and Figure 4 shows the probability for no change for speech and silence frames respectively.

III. CHANNEL DECODING WITH APRI-SOVA

We have considered the transmission model shown in Figure 5 consisting of source encoder, channel encoder, AWGN-channel, channel decoder and source decoder. The Soft Output Viterbi Algorithm (SOVA) introduced by Hag enauer [1], allows to obtain a reli-

---

**Fig. 1.** GSM-AMR channel coding (a) for speech frames (b) for updated silence frames

**Table 1**

<table>
<thead>
<tr>
<th>Mode</th>
<th>coded bits</th>
<th>class1a</th>
<th>rate</th>
<th>convolutional coder memory, l</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.2</td>
<td>244</td>
<td>81</td>
<td>1/2</td>
<td>4</td>
</tr>
<tr>
<td>10.2</td>
<td>204</td>
<td>65</td>
<td>1/3</td>
<td>4</td>
</tr>
<tr>
<td>7.95</td>
<td>159</td>
<td>75</td>
<td>1/3</td>
<td>6</td>
</tr>
<tr>
<td>7.4</td>
<td>148</td>
<td>61</td>
<td>1/3</td>
<td>4</td>
</tr>
<tr>
<td>6.7</td>
<td>134</td>
<td>55</td>
<td>1/4</td>
<td>4</td>
</tr>
<tr>
<td>5.9</td>
<td>118</td>
<td>55</td>
<td>1/4</td>
<td>6</td>
</tr>
<tr>
<td>5.15</td>
<td>103</td>
<td>49</td>
<td>1/5</td>
<td>4</td>
</tr>
<tr>
<td>4.75</td>
<td>95</td>
<td>39</td>
<td>1/5</td>
<td>6</td>
</tr>
</tbody>
</table>

The output of the channel decoder contains bit reliability information that is used for error recovery at the source decoder. The Soft Output Viterbi Algorithm (SOVA) introduces soft output for the decoder branch which is defined as the probability of each branch being correct. The SOVA algorithm is used to obtain the soft output bits for the decoder branch which is defined as the probability of each branch being correct. The SOVA algorithm is used to obtain the soft output bits for the decoder branch which is defined as the probability of each branch being correct.
ability information $L(\hat{u}_i) = \log \left[ \frac{p(\hat{u}_i=+1)}{p(\hat{u}_i=-1)} \right]$ of the decoded information bits $\hat{u}_i$, with only few changes [4] to the Viterbi Algorithm. In addition to the channel values $y_i$, the APRI-SOV A considers a priori information $L(u_i) = \log \left[ \frac{p(u_i=+1)}{p(u_i=-1)} \right]$ on the source bits $u_i$. This a priori information leads to significantly fewer bit errors compared to the Viterbi Algorithm, especially in a low signal-to-noise ratio environment. The new metric $M_i(m)$ of path $m_i$ for a binary trellis labeled with $N$ coded bits $x_{i,n}$ and the information bits $u_i$ as,

$$M_i^{(m)} = M_i^{(m)} + \sum_{n=1}^{N} x_{i,n} \cdot L_{c,i,n} \cdot y_{i,n} + u_i^{m} \cdot L(u_i) \quad (1)$$

where

$$L_{c} \cdot y = \log \left[ \frac{p(y/x = +1)}{p(y/x = -1)} \right], \quad L_{c} = 4 \cdot \frac{E_i}{N_o} \quad (2)$$

We have implemented a simple algorithm proposed by Hagenauer [1] to build $L(u)$ from the soft output of previous frame $L(\hat{u}_{k-1})$ and the estimated bit chang-

Fig. 2. Interframe correlation

ability probability $L(c)$, where $c$ is a change bit,

$$L(u_{k,q}) = L(\hat{u}_{k-1,q}) \oplus L(\hat{u}_{k,q})$$

$$= L(\hat{u}_{k-1,q} \oplus c_{k,q})$$

$$= \log \left[ 1 + e^{L(\hat{u}_{k-1,q}) \cdot c_{k,q}} + e^{L(\hat{u}_{k,q})} \right] \quad (3)$$

For this we have used the following approximation [1]

$$L(u_{k,q}) \approx \text{sign}(L(\hat{u}_{k-1,q}) \cdot \text{sign}(L(\hat{u}_{k,q})) \cdot \min(|\hat{u}_{k-1,q}|, |\hat{u}_{k,q}|) \quad (4)$$

Since speech is highly non-stationary the change bit reliability values $L(c)$ have to be calculated for every frame. We have used the model proposed by Hindelang [2] for calculating $L(c)$ which is shown in Figure 6.

III-A. APRI-SOV A with VAD

While operating with VAD, the GSM-AMR codec segregates the incoming source data into speech and
state | P(c) | L(c)  
---|---|---  
s15 | 0.95 | 3.0  
s14 |  
s3  |  
s2  |  
s1  | 0.55 | 0.2  
s0  | 0.5  | 0  

Fig. 6. Modified HUK model

Fig. 7. Application of APRI-SOVA to GSM-AMR channel decoder

silence frames. The channel coder needs to handle the speech and silence frames separately. Output of the speech coder contains blocks of speech and silence frames alternately. APRI-SOVA requires old decoded frame reliability values for decoding new frame. While applying the SCCD using APRI-SOVA, blocks of silence frames can be considered to be continuous as the parameters of background noise remain almost constant. Hence we retain the reliability values of last decoded silence frame for decoding the new silence frame. The same method can be used for speech frames also, but parameters of two speech frames separated by silence frames may differ considerably. So, the reliability values of last decoded speech frame may not be useful. We have observed that the difference in performance when using reset and retained reliability values is negligible. We choose to retain the reliability values of last decoded frames both in silence and speech cases.

IV. SIMULATION RESULTS

We have taken a speech file from NTIMIT data base. After the speech encoding [5] of this file, there are 25000 speech frames and 10000 silence frames. We have observed an interframe correlation both in speech and silence frames. We have estimated the probability for change $P_c$ for all bits in silence and speech frames. For example in speech frame of 12.2 kbps mode, the bits with indices 1, 8, 10, 30, 31, 32, 33, 50, 51, 60, 61, 62 and 63 have the $P_c$ value less or equal to 0.3. In silence frames the bits with indices 1 to 23 and 26 to 34 have the $P_c$ value less or equal to 0.3. Figure 3 shows probability for no change $1 - P_c$, in speech frames of 12.2 kbps mode. Figure 4 shows probability for no change $1 - P_c$, in silence frames.

We have considered the model for transmission shown in section 3 for AWGN channel, and we have applied SCCD with APRI-SOVA to those bits with indices whose $P_c$ value is less than or equal to 0.3 for speech and silence frames in all modes. We have compared the BER values at different SNRs with classical Viterbi Algorithm. Figure 8 shows performance comparison of APRI-SOVA over VA for speech frames in 10.2 kbps mode by disabling VAD, in terms of BER. Figure 9 shows performance comparison of APRI-SOVA over VA for speech frames in 7.4 kbps mode by enabling VAD, in terms of BER. Figure 10 shows performance comparison of APRI-SOVA over VA for silence frames in terms of BER.

V. CONCLUSIONS

We have applied SCCD using APRI-SOVA to the GSM-AMR system with and without enabling VAD. We have observed a gain of 0.5 dB when VAD is disabled for 10.2 kbps mode and a gain of 0.4 dB when VAD is enabled for 7.4 kbps mode in terms of $E_b/N_0$ over classical VA. We have observed a gradual decrease in gain as bitrate goes down from 12.2 kbps to 4.75 kbps.

For silence frames, due to their high interframe correlation, APRI-SOVA is expected to perform much bet-
ter than VA but simulations indicate that the performance difference is only marginal. This may be because VA itself handles silence frames quite well so that the gain for APRI-SOV A is not very high.

All simulations have been done for the AWGN channel model. Presently we are extending our work to different fading channel models.

Fig. 8. performance comparison of APRI-SOV A over VA for speech frames in 10.2kbps mode of GSM-AMR system by disabling VAD

Fig. 9. performance comparison of APRI-SOV A over VA for speech frames in 7.4kbps mode of GSM-AMR system by enabling VAD

Fig. 10. performance comparison of APRI-SOV A over VA for silence frames of GSM-AMR system

REFERENCES


