ON DETERMINING SOFT OUTPUT OF THE CELLULAR TEXT TELEPHONE MODEM (CTM) DEMODULATOR

Wen Xu and Matthias Marke

Department of Mobile Phone Development, Siemens AG,
Grillparzerstr. 10 - 18, D-81675 Munich, Germany

ABSTRACT

Using soft output of a demodulator is essential to improve performance of the following channel decoder, and performance of the overall communication system. In this study, aspects on determining the soft output of the Cellular Text telephone Modem (CTM) demodulator are described. The optimal soft value of individual demodulated bits, which is explicitly related to correlator output for symbols, is derived for an AWGN channel. When additional information from the cellular receiver, such as the bad (speech) frame indicator, is available at the CTM receiver, it can be used to properly manipulate the obtained soft output of the CTM demodulator. By employing the CTM receiver optimized in this way for text transmission via the GSM-based PCS1900, a performance gain of up to 1 - 2 dB in channel signal-to-noise ratio can be obtained, in comparison with the example solution by the 3rd Generation Partnership Project (3GPP).

1. INTRODUCTION

Specific Text Telephony (referred to as TTY in North America, see the recommendations ITU-T V.18 and T.140) equipments have been used in the fixed network for many years, especially by deaf, hard of hearing and speech-impaired persons, to transmit text and speech through ordinary speech traffic channels. The FCC under the US government also requires that transmission of the emergency calls (911) be guaranteed in cellular mobile systems. For these reasons, the 3rd Generation Partnership Project (3GPP) recently standardized the so-called Cellular Text Telephone Modem (CTM) [1, 5], a solution for the GSM, UMTS and potentially other cellular technologies. The CTM aims at reliable transmission of a text telephone conversation at a speed of 80 net bits per second (i.e. 10 char/s), alternating with a speech conversation through the existing speech communication paths in cellular mobile systems. Together with TTY, CTM may serve for worldwide applications in text telephony.

As depicted in Fig. 1, the CTM transmitter expects text input in ISO10646-1 encoding. It performs character encoding in UTF-8 format, forward error correction (FEC) with a memory = 4, rate = 1/4 convolutional code, periodical muting and insertion of synchronization bits, interleaving and 4-FSK modulation. Each pair of two bits (= a symbol) from the interleaver is modulated into a sine waveform of a length of 5 ms (40 samples) starting with a phase of zero. Depending on the values of the two bits, an artificial tone signal with one of the frequencies 400, 600, 800, and 1000 Hz is generated, and sent, like a speech signal, to the cellular transmitter for speech encoding, channel encoding, etc. In case there is no text to transmit, the switch S1 is set to the ‘closed’ position so that a speech or audio signal can bypass the CTM transmitter.

At the receiving end, the CTM signal is detected and decoded correspondingly, resulting in the “text” output. If the CTM demodulator does not detect a CTM signal, the speech or audio signal coming from the speech decoder is forwarded via the switch S2 to the “speech” output.

As can be seen, the text is transmitted through a cascaded system of the CTM and the cellular system. The 3GPP CTM solution has been intensively optimized during the standardization. However, especially due to the following reasons, the CTM and cellular system do not match to each other and the cascaded system may not provide enough high performance:

a) The speech encoder, such as the AMR (adaptive multi-rate) coder used in GSM and UMTS [3], is optimized for compression of the human speeches, not the artificial tone signals.

b) The error concealment algorithm, usually embedded in the speech decoder and enabled when speech frame is heavily degraded, is optimized for human ears and may not be appropriate for transmission of text information.

0-7803-7663-3/03/$17.00 ©2003 IEEE

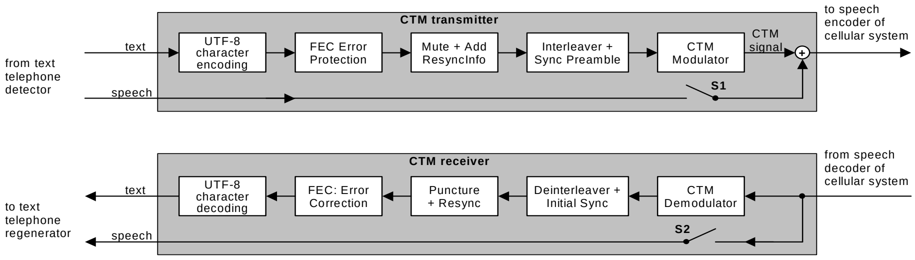

Fig. 1. Overview of the CTM transmitter and receiver.

In this study, aspects on determining the soft output of the CTM demodulator are described. The basic idea is to optimize soft input/soft output in the CTM receiver for a better overall performance. The aspects have been successfully employed in text transmission via the CTM and the GSM (here the American PCS1900). Since the optimization is done in the CTM receiver only, the specified standard remains untouched.

2. SOFT VALUE CALCULATION IN CTM DEMODULATOR

Soft output of a demodulator, which represents the reliability of the demodulated bits, serves as soft input for channel decoder which follows. Using the (optimal) soft output of the demodulator is essential to improve performance of the channel decoder, and performance of the overall communication system. For a cascaded system, it is in general important to use the optimal soft input and soft output, whenever possible, in all receiver stages in order to achieve good performance of the overall system [6]. This is especially true for the CTM (see Fig. 1).

In the 3GPP example solution, reliabilities (= magnitudes of the soft output) of the CTM-demodulated two bits of a symbol are assumed to be the same. They are calculated by subtracting the mean of the three cross correlations of the not chosen frequency from the cross correlation of the chosen frequency [2]. This is a sub-optimal solution, since in general, the reliabilities of individual bits of a symbol are not the same.

In what follows, we derive, under the assumption of an AWGN channel, the optimal soft outputs of individual bits of a symbol in the maximum *a posteriori* probability (MAP) or maximum likelihood (ML) demodulation based on a correlator or matched-filter.

For simplicity, let us denote the (modulated) sent signal sequence to be $\vec{x}^{(m)} = (x_0^{(m)}, x_1^{(m)}, \ldots, x_{L-1}^{(m)})$, where $m \in$ {0, 1, …, M–1} is the index of possible sent signal sequences; L is the number of the considered data samples. The corresponding state sequence (symbol) of $\vec{x}^{(m)}$ is $\vec{s}^{(m)} = (s_{K-1}, \ldots, s_1, s_0)$, where $\vec{s}^{(m)}$ is the binary representation of m with K bits (i.e. $s_i \in \{+1, -1\}$, i = 0, 1, …, K–1, $2^K \geq M$) and thus has a one-to-one relation to $\vec{x}^{(m)}$. The received sequence is $\vec{y} = (y_0, y_1, \ldots, y_{L-1})$.

A maximum *a posteriori* probability (MAP) detector selects $\vec{x}^{(m)}$ which maximizes the *a posteriori* probability $P(\vec{x}^{(m)} \mid \vec{y})$. A maximum likelihood (ML) detector selects $\vec{x}^{(m)}$ which maximizes the likelihood function $P(\vec{y} \mid \vec{x}^{(m)})$. An MAP detector and an ML detector can be related with each other through

$$P(\vec{x}^{(m)} \mid \vec{y}) = \alpha_m P(\vec{y} \mid \vec{x}^{(m)}) \tag{1}$$

where

$$\alpha_m \stackrel{\Delta}{=} \frac{P(\vec{x}^{(m)})}{P(\vec{y})} \tag{2}$$

$P(\vec{x}^{(m)})$ is the *a priori* information of $\vec{x}^{(m)}$.

In what follows, we derive the soft value of s_i, which is defined as

$$L(s_i) \stackrel{\Delta}{=} \log \frac{P(s_i = +1 \mid \vec{y})}{P(s_i = -1 \mid \vec{y})} \tag{3}$$

The transmission of $\vec{x}^{(m)}$ over an AWGN channel is modeled as

$$\vec{y} = \vec{x}^{(m)} + \vec{n} \tag{4}$$

where $\vec{n}$ is the white Gaussian noise vector (with zero mean). It holds then

$$\begin{aligned} P(\vec{y} \mid \vec{x}^{(m)}) &= \prod_{j=0}^{L-1} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_j - x_j^{(m)})^2}{2\sigma^2} \right) \\ &= \frac{1}{\left(\sqrt{2\pi}\sigma\right)^L} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=0}^{L-1} (y_j{}^2 - 2 y_j x_j^{(m)} + x_j^{(m)2}) \right) \end{aligned} \tag{5}$$

where σ is the standard deviation of y_j and can be estimated as

$$\begin{aligned} \bar{y} &\approx \frac{1}{L} \sum_{j=0}^{L-1} y_j \\ \sigma &\approx \sqrt{\frac{1}{L} \sum_{j=0}^{L-1} (y_j - \bar{y})^2} \end{aligned} \tag{6}$$

According to Eq. (4), we have for $\vec{x}^{(m)}$ ($m \in$ {0, 1, …, M–1}) having the same energy

$$\frac{1}{2\sigma^2}\sum_{j=0}^{L-1}(y_j^{\,2}+x_j^{(m)\,2})\overset{\Delta}{=}\beta=const \tag{7}$$

In addition, we define

$$\rho^{(m)}\overset{\Delta}{=}corr(\vec{y},\vec{x}^{(m)})=\sum_{j=0}^{L-1}y_j x_j^{(m)} \tag{8}$$

i.e. $\rho^{(m)}$ is the output of a correlation receiver or a matched filter, then

$$P(\vec{y}\mid\vec{x}^{(m)})=\frac{1}{\left(\sqrt{2\pi}\sigma\right)^L}\exp\left(\rho^{(m)}/\sigma^2-\beta\right) \tag{9}$$

Due to the one-to-one relation between $\vec{x}^{(m)}$ and $\vec{s}^{(m)}$, we have

$$P(\vec{s}^{(m)}\mid\vec{y})=P(\vec{x}^{(m)}\mid\vec{y})=\frac{\alpha_m}{\left(\sqrt{2\pi}\sigma\right)^L}\exp\left(\rho^{(m)}/\sigma^2-\beta\right) \tag{10}$$

Since

$$P(s_i=+1\mid\vec{y})=\sum_{m,where\, s_i=+1}P(\vec{s}^{(m)}\mid\vec{y})$$

$$P(s_i=-1\mid\vec{y})=\sum_{m,where\, s_i=-1}P(\vec{s}^{(m)}\mid\vec{y})$$

the soft output for an MAP detection becomes therefore

$$L(s_i)=\log\frac{\sum_{m,where\, s_i=+1}P(\vec{s}^{(m)}\mid\vec{y})}{\sum_{m,where\, s_i=-1}P(\vec{s}^{(m)}\mid\vec{y})} \tag{11}$$

or as a reasonable approximation

$$L(s_i)\approx\log\frac{\max_{m,where\, s_i=+1}P(\vec{s}^{(m)}\mid\vec{y})}{\max_{m,where\, s_i=-1}P(\vec{s}^{(m)}\mid\vec{y})}\overset{\Delta}{=}\tilde{L}(s_i) \tag{12}$$

Using Eq. (10), we have

$$\begin{aligned}L(s_i)&=\log\frac{\sum_{m,where\, s_i=+1}\alpha_m\exp(\rho^{(m)}/\sigma^2)}{\sum_{m,where\, s_i=-1}\alpha_m\exp(\rho^{(m)}/\sigma^2)}\\&\approx\max_{m,where\, s_i=+1}(\rho^{(m)}/\sigma^2+\log\alpha_m)\\&-\max_{m,where\, s_i=-1}(\rho^{(m)}/\sigma^2+\log\alpha_m)=\tilde{L}(s_i)\end{aligned} \tag{13}$$

Now denote M^+ be the number of the indices m, where $s_i = +1$, and M^- be the number of the indices m, where $s_i = -1$. Then we have

$$\begin{aligned}L(s_i)&\geq\log\frac{\sum_{m,where\, s_i=+1}P(\vec{s}^{(m)}\mid\vec{y})}{M^-\max_{m,where\, s_i=-1}P(\vec{s}^{(m)}\mid\vec{y})}\\&\geq\log\frac{\max_{m,where\, s_i=+1}P(\vec{s}^{(m)}\mid\vec{y})}{M^-\max_{m,where\, s_i=-1}P(\vec{s}^{(m)}\mid\vec{y})}\end{aligned} \tag{14}$$

$$L(s_i)\leq\log\frac{M^+\max_{m,where\, s_i=+1}P(\vec{s}^{(m)}\mid\vec{y})}{\max_{m,where\, s_i=-1}P(\vec{s}^{(m)}\mid\vec{y})} \tag{15}$$

$L(s_i)$ is therefore upper- and lower-bounded by

$$\tilde{L}(s_i)-\log M^-\leq L(s_i)\leq\tilde{L}(s_i)+\log M^+ \tag{16}$$

We see that Eq. (12) provides a good approximation.

In the case that no *a priori* information $P(\vec{x}^{(m)})$ is available or $P(\vec{x}^{(m)})$ is constant, the MAP detector is equivalent to the ML detector and Eq. (13) reduces to

$$\begin{aligned}L(s_i)&=\log\frac{\sum_{m,where\, s_i=+1}\exp(\rho^{(m)}/\sigma^2)}{\sum_{m,where\, s_i=-1}\exp(\rho^{(m)}/\sigma^2)}\\&\approx\frac{1}{\sigma^2}\left(\max_{m,where\, s_i=+1}\rho^{(m)}-\max_{m,where\, s_i=-1}\rho^{(m)}\right)\end{aligned} \tag{17}$$

The approximation in the above equations (12), (13) and (17) can greatly simplify the implementation and provide adequate accuracy in many applications, such as in the CTM demodulation. Due to the optimality of the method to calculate the soft output, significant performance gain can be obtained.

In the case of the CTM, a pair of two bits $\vec{s}^{(m)}$ = (s_1, s_0) are modulated using a 4-ary FSK into one of the four sine waveforms (400, 600, 800 and 1000 Hz) of a length of 5 ms $\vec{x}^{(m)}=(x_0^{(m)},x_1^{(m)},\ ...,x_{L-1}^{(m)})$ (L = 40 samples, m = 0, 1, 2, 3). With "+1" representing the logical 1, "–1" the logical 0, the mapping between a bit pair and a sine frequency can be written as

$$\begin{aligned}\vec{s}^{(0)}&=(-1,-1)\\\vec{s}^{(1)}&=(-1,+1)\\\vec{s}^{(2)}&=(+1,-1)\\\vec{s}^{(3)}&=(+1,+1)\end{aligned} \tag{18}$$

By omitting the *a priori* information $P(\vec{x}^{(m)})$, we have

$$\begin{aligned}L(s_0)&\approx\frac{1}{\sigma^2}\left(\max(\rho^{(1)},\rho^{(3)})-\max(\rho^{(0)},\rho^{(2)})\right)\\L(s_1)&\approx\frac{1}{\sigma^2}\left(\max(\rho^{(2)},\rho^{(3)})-\max(\rho^{(0)},\rho^{(1)})\right)\end{aligned} \tag{19}$$

where the cross-correlation $\rho^{(m)}$ may be calculated after the received samples are properly low-pass filtered.

Our simulations showed that for the CTM, almost the same performance can be achieved when the simple approximation in Eq. (12) instead of the exact formula in Eq. (11) is used to compute the soft value. σ^2 can be chosen such that best performance is achieved.

For comparison the reliability of the bits (s_1, s_0) in the 3GPP example solution [2] is given by

$$|L(s_0)| = |L(s_1)| = k_0\left(\rho^{(m0)} - \frac{1}{3}\sum_{m\neq m0}\rho^{(m)}\right) \quad (20)$$

where $m0$ is the index m for which $\rho^{(m)}$ achieves the maximum, and k_0 is a constant. The sign of $L(s_i)$ is determined by the mapping between a bit pair and a sine frequency, i.e. Eq. (18).

3. APPLICATION IN TEXT TRANSMISSION VIA CTM AND PCS1900

PCS1900 cellular phone system is a GSM variant deployed in the US. It operates at 1900 MHz frequency band. The AMR speech codec has been standardized as the mandatory speech codec for the GSM and UMTS [3], and is introduced in the PCS1900 as well. It operates in 8 modes, namely 4.75, 5.15, 5.90, 6.70, 7.40, 7.90, 10.2, 12.2 kbit/s.

With the AMR used in the GSM, an FEC follows the AMR encoded speeches. Very important bits of the AMR encoded speech parameters, the so-called class 1a bits, are protected by a CRC parity check. If the parity check at the receiving side fails, the speech frame is marked with a bad frame indicator (BFI).

This BFI is submitted to the AMR speech decoding module to control the error concealment mechanism, as specified in [4]. Purpose of the error concealment is to conceal or avoid possible annoying effects due to the loss of speech frames. Its basic idea is to artificially generate a "pleasant" speech frame based on available "good" speech frames, by repeating or extrapolating speech parameters of previous frames

Using the speech channel for transmission of text data, the error concealment may result in a negative effect in transmission quality. A bad frame may indicate a bad speech quality but may still contain enough information to restore the CTM signal. This information will be lost when parameters of current frame are replaced by those of previous frames to construct an audio signal which sounds pleasant to listeners. This has consequently a negative impact on performance of the following CTM channel decoding.

To counteract this effect, we used the BFI, i.e. the information whether an AMR speech frame is bad or good, to manipulate (weight) the soft output of the CTM demodulator. Concretely, the soft outputs of the CTM demodulator will be properly attenuated when they belong to bad frames. This, according to our investigations, leads to further performance gains.

We carried out simulations for text message transmission via a GMSK-modulated PCS1900 AMR speech channel under typical mobile channel conditions. A long message of random characters corresponding to about 500000 speech frames (= 10000 seconds) was used as CTM input text. The decoded text message was analyzed by determining the number of character errors that have been caused due to the transmission. The character error rate (CER) is defined as the ratio of the number of all errors divided by number of the original text, where the number of all errors is the sum of the number of deleted characters, the number of inserted characters and the number of replaced characters.

We improved the 3GPP example solution by using the methods described above. With the method to optimally calculate the soft value of the CTM-demodulated bits alone (e.g. Eq. (19)), we have a typical gain of about 0.5 - 1.5 dB in channel signal-to-noise ratio (SNR), for the same CER. Notice that the performance gains obtained slightly depend on the AMR mode used.

In addition, if the BFI is available at the CTM receiver, it can be used to properly weight the soft output optimally calculated (e.g. according to Eq. (19)) for further performance improvement. In this case, a total gain of up to 1.0 - 2.0 dB in channel SNR can be obtained. This corresponds to, for the same channel SNR, a reduction of CER by a factor of about 3 - 5.

4. CONCLUSIONS

In this study, we investigated some aspects on reliable text transmission through the CTM. The optimal soft value of individual demodulated bits is derived for the AWGN channel. It is explicitly related to correlator (or matched-filter) output for symbols. Additional information from the cellular receiver, such as the bad (speech) frame indicator, can be used to properly manipulate the soft output of the CTM demodulator. By employing the CTM receiver optimized in this way in text transmission via the GSM-based PCS1900, a performance gain of 1 - 2 dB in channel signal-to-noise ratio can be obtained in comparison with the 3GPP example solution.

5. REFERENCES

[1] 3GPP TS 26.226, "Cellular Text Telephone Modem (CTM), General Description (Rel. 5)," 2001.

[2] 3GPP TS 26.230, "Cellular Text Telephone Modem (CTM), Transmitter Bit Exact C-Code (Rel. 5)," 2001.

[3] 3GPP TS 26.090, "AMR speech codec; Transcoding functions (Rel. 4)," 2001.

[4] 3GPP TS 26.091, "AMR speech codec; Error concealment of lost frames (Rel. 4)," 2001.

[5] M. Dörbecker, K. Hellwig, F. Jansson and T. Frankkila, "The cellular text telephone modem - the solution for supporting text telephone functionality in GSM networks," in *Proc. ICASSP'01*, pp. 1441-1444, 2001.

[6] J. Hagenauer, "Soft-In/Soft-Out: The benefits of using soft values in all stages of digital receivers," in *Proc. of 3rd Int. Workshop on DSP Techniques Applied to Space Comm.*, ESTEC, Noordwijk, The Netherlands, Sept. 1992.