A PROBABILISTIC FRAMEWORK FOR
OPTIMUM SPEECH EXTRAPOLATION IN DIGITAL MOBILE RADIO

Christian G. Gerlach

Institute for Communication Systems and Data Processing
Aachen University of Technology
Templergraben 55, W-5100 Aachen, Germany

ABSTRACT

In digital mobile radio systems the speech quality can be degraded severly if the channel decoder produces residual bit errors, e. g. due to heavy burst errors on the channel. A new combined speech extrapolation and error detection algorithm is presented which mostly improves the speech significantly in case of residual bit errors. This algorithm, which is part of the speech decoding process, uses a posteriori–probabilities of speech parameters.

1. INTRODUCTION

In digital mobile radio systems speech coding techniques are applied to reduce the transmission bit rate and channel coding techniques are used for error protection. A typical example is the pan-European Digital Mobile Radio System "GSM" with a net bit rate of 13 kbps for speech transmission and 9.8 kbps for error protection, i. e. with a gross bit rate of 22.8 kbps [1, 2], see Fig. 1.

In certain situations, especially in case of severe burst errors, the channel decoding scheme may be "overloaded" resulting in residual bit errors and in severe degradation of speech quality. However, the speech quality can be improved by a repairing process, called speech extrapolation, which is not applied to individual bits but to individual parameters of the speech codec.

Usually speech coding is based on the more or less independent processing of consecutive frames of speech samples. For this reason there is still an inherent correlation between consecutive frames which can be exploited.

In the GSM system a simple speech extrapolation technique is applied by repeating the parameters of the previous 20 ms frame if residual errors are detected (bad frame indication). A similar approach is used in the TDMA proposal IS-54 for the digital cellular system in North America [3]. If half-rate codecs are introduced, the need for good speech extrapolation is even increased.

Until now error detection and speech extrapolation are treated as separate problems e. g. [3, 4]. In this contribution a new probabilistic framework for combined speech extrapolation and error detection will be developed. Based on the *a posteriori*-probability of the coder parameters which can be computed, if the previous parameters were received correctly. With this *a posteriori*-probability optimum estimators adapted to human perception can be applied and soft decision information can fully be exploited.

2. SOURCE PARAMETER REDUNDANCY

For simplicity the principle of the new extrapolation scheme will be explained by considering samples $\gamma(k)$
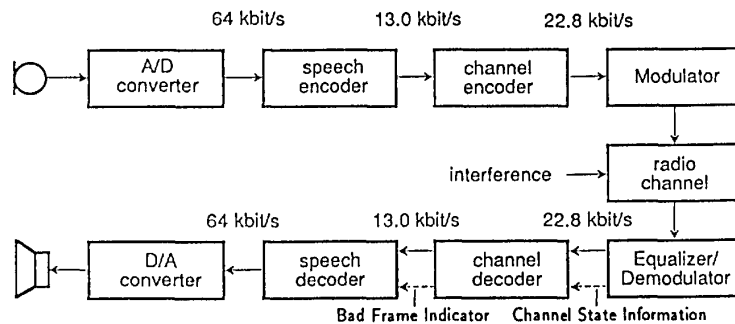

Figure 1: Speech transmission in the European digital mobile radio system (GSM)

of a single codec parameter, e. g. a LPC-coefficient or a gain factor. With k denoting time, e. g. frame or subframe index, the sequence $\gamma(k)$ is transmitted. Our knowledge about the statistical dependency of consecutive parameters is covered by the conditional probability density at time k given all previous parameters $\gamma(k-1), \gamma(k-2), \ldots, \gamma(k-M)$

$$p_{\gamma(k)|\gamma(k-1),\gamma(k-2),\ldots,\gamma(k-M)}(\xi|\eta_1, \eta_2, \ldots, \eta_M) . \quad (1)$$

We suppose that this function is known. If $\gamma(k)$ can be modeled e. g. by a coloured Gaussian process with Markoff property of M-th order, then the conditional probability is given by

$$p_{\gamma(k)|\hat\gamma(k)}(\xi|\hat\gamma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\xi-\hat\gamma)^2}{2\sigma^2}} \quad (2)$$

with

$$\hat\gamma(k) = E\{\gamma(k)|\gamma(k-1), \ldots, \gamma(k-M)\} = \sum_{i=1}^{M} h_i \, \gamma(k-i) \quad (3)$$

the conditional mean that can be obtained by linear prediction with coefficients h_i. The quantized version of γ is denoted by γ_q taking 2^m discrete, possibly nonuniform levels. They are represented by m bits $a_1, a_2, \ldots, a_m$ according to the coding law $a_i = q_i(\gamma_q)$.

Now the conditional probability of the quantized value $\gamma_q(k)$ or the corresponding bit combination $a_1, a_2, \ldots, a_m$ is obtained by integrating $p_{\gamma|\hat\gamma}$ over the quantization intervals as indicated in Fig. 2. This always results in the
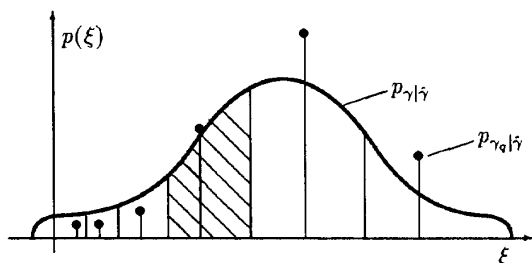
Figure 2: Relation between the distribution of γ and γ_q

discrete distribution

$$p_{\gamma_q(k)|\gamma(k-1),\ldots,\gamma(k-M)}(\xi|\eta_1, \ldots, \eta_M) = p_{\gamma_q(k)}(\xi) \quad (4)$$

for $\gamma_q(k)$ or the corresponding bit combination. If the receiver gets the previous samples $\gamma(k-1), \ldots, \gamma(k-M)$ correctly, the one dimensional function $p_{\gamma_q(k)}(\xi)$ can be computed. This function describes the a priori-probability for the next quantized sample γ_q. For simplicity the time index k will be omitted in the sequel.

3. TRANSMISSION MODEL

3.1 Soft Bit Considerations

For the transmission of the individual bits the following simple model is applied, see Fig. 3a. For simplicity we assume "bipolar bits" $a_i \in \{-1, 1\}$. The transmission is
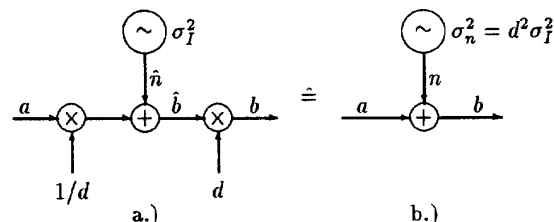
Figure 3: Transmission model for a single bit.

modeled by an attenuation $d > 1$ and by superposition of gaussian noise n, which is statistical independent of a. It can be shown that this model is valid e. g. for MSK with matched filter detection and that it can be used approximately for GMSK too.

Due to the carrier amplitude measurement the attenuation factor d is approximately known, and the soft decision information $sd = \frac{2\hat b}{d\sigma_I^2}$ is available at the receiver. With $b = d \cdot \hat b$ the model Fig. 3b can be used equivalently. The amplified output value b is described by

$$p_{b|a}(y|x) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{(y-x)^2}{2\sigma_n^2}} \quad (5)$$

with the variance of the effective noise $\sigma_n^2 = d^2\sigma_I^2$.

If $sd = \frac{2\hat b}{d\sigma_I^2} = \frac{2b}{\sigma_n^2}$ is given, the probability of a wrong threshold decision can be computed as shown below. The a priori-probability of a is given by the discrete distribution $\tilde p_a(x) = p_a(-1)\delta(x+1) + p_a(1)\delta(x-1)$, and the a posteriori-probability can be derived as

$$p_{a|b}(x|y) = \frac{p_a(-1)\delta(x+1) + p_a(1)\delta(x-1)e^{\frac{2y}{\sigma_n^2}}}{p_a(-1) + p_a(1)e^{\frac{2y}{\sigma_n^2}}} \quad (6)$$

Assuming symmetric a priori-probabilities $p_a(1) = p_a(-1) = 0.5$ we get

$$p_{a|b}(x|y) = \frac{\delta(x+1) + \delta(x-1)e^{\frac{2y}{\sigma_n^2}}}{1 + e^{\frac{2y}{\sigma_n^2}}} . \quad (7)$$

If the decoder decides in case of $b > 0$, that $a = 1$ was sent, the error probability $P_e = p_{a|b}(-1|y = b)$ is given by

$$P_e = \frac{1}{1 + e^{\frac{2b}{\sigma_n^2}}} = \frac{1}{1 + e^{sd}} \quad (8)$$

leading to the relation for the soft decision information

$$\frac{2\hat{b}}{d\sigma_f^2} = \frac{2b}{\sigma_n^2} = \text{sd} = -\ln\left(\frac{P_e}{1-P_e}\right). \quad (9)$$

Hence, the error mechanism can be completely described using (5). If P_e is derived from the channel decoder equation (9) provides an alternative approach to obtain sd. If the channel decoder does not produce soft decision output in contrast to [5], mean transition probabilities must be used instead.

3.2 Transmission of a Bit Combination

With the model illustrated in Fig. 3b the transmission can be described by Fig. 4. For the received $b_i, i \in \{1, \ldots, m\}$
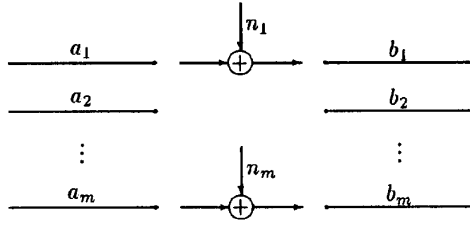

Figure 4: Transmission model for a bit combination.

we obtain

$$p_{b_i|a_i}(y|x) = \frac{1}{\sqrt{2\pi\sigma_{n_i}^2}} e^{-\frac{(y-x)^2}{2\sigma_{n_i}^2}}. \quad (10)$$

Due to bitshuffling, interleaving and equalization the disturbance of the bits is approximately statistical independent. Hence, the conditional joint distribution function is the product of marginal distributions:

$$p_{b_1,\ldots,b_m|a_1,\ldots,a_m}(y_1,\ldots,y_m|x_1,\ldots,x_m)$$

$$= \prod_{i=1}^{m} p_{b_i|a_i}(y_i|x_i)$$

$$= \frac{1}{\sqrt{2\pi\sigma_{n_1}^2}} \cdots \frac{1}{\sqrt{2\pi\sigma_{n_m}^2}} e^{-\frac{1}{2}\left[\frac{(y_1-x_1)^2}{\sigma_{n_1}^2}+\cdots+\frac{(y_m-x_m)^2}{\sigma_{n_m}^2}\right]}. \quad (11)$$

Considering that every a_i is a function of γ_q (coding law) leads to the transition probability for given γ_q:

$$p_{b_1,\ldots,b_m|\gamma_q}(y_1,\ldots,y_m|\xi)$$

$$= p_{b_1,\ldots,b_m|a_1,\ldots,a_l}(y_1,\ldots,y_m|q_1(\xi),\ldots,q_m(\xi))$$

$$= \frac{1}{\sqrt{2\pi\sigma_{n_1}^2}} \cdots \frac{1}{\sqrt{2\pi\sigma_{n_m}^2}} \quad (12)$$

$$\cdot e^{-\frac{1}{2}\left[\frac{(y_1-q_1(\xi))^2}{\sigma_{n_1}^2}+\cdots+\frac{(y_m-q_m(\xi))^2}{\sigma_{n_m}^2}\right]}.$$

4. NEW APPROACH

Substituting $y_1, \ldots, y_m$ by the received values $b_1, \ldots, b_m$, the quantized speech parameter γ_q can be estimated by maximizing (12) versus ξ. This is the conventional maximum likelihood decoding approach, efficiently implemented by (soft decision) Viterbi decoding.

In the new approach, however, additionally the *a priori*-probability (4) is considered and the joint distribution is determined as

$$p_{b_1,\ldots,b_m,\gamma_q}(y_1,\ldots,y_m,\xi)$$

$$= p_{b_1,\ldots,b_m|\gamma_q}(y_1,\ldots,y_m|\xi) \cdot p_{\gamma_q}(\xi). \quad (13)$$

Equation (13) represents complete statistical information about the transmission system. Based on (13) the *a posteriori*-probability of the quantized speech parameter γ_q can be calculated, using summation over 2^m quantization levels:

$$p_{\gamma_q|b_1,\ldots,b_m}(\xi|y_1,\ldots,y_m) = \frac{p_{b_1,\ldots,b_m,\gamma_q}(y_1,\ldots,y_m,\xi)}{p_{b_1,\ldots,b_m}(y_1,\ldots,y_m)}$$

$$= \frac{p_{b_1,\ldots,b_m,\gamma_q}(y_1,\ldots,y_m,\xi)}{\sum\limits_{j=1}^{2^m} p_{b_1,\ldots,b_m,\gamma_q}(y_1,\ldots,y_m,\xi_j)}. \quad (14)$$

It can be shown after calculation that the following relation is valid

$$p_{\gamma_q|b_1,\ldots,b_m}(\xi|y_1,\ldots,y_m)$$

$$= \frac{e^{+\left[\frac{y_1}{\sigma_{n_1}^2}\cdot q_1(\xi)+\cdots+\frac{y_m}{\sigma_{n_m}^2}\cdot q_m(\xi)\right]} \cdot p_{\gamma_q}(\xi)}{\sum\limits_{j=1}^{2^m} p_{\gamma_q}(\xi_j)\cdot e^{+\left[\frac{y_1}{\sigma_{n_1}^2}\cdot q_1(\xi_j)+\cdots+\frac{y_m}{\sigma_{n_m}^2}\cdot q_m(\xi_j)\right]}}, \quad (15)$$

taking into consideration the identity $q_i^2(\xi) = 1$. With the received soft decision bits $\frac{y_i}{\sigma_{n_i}^2} = \frac{sd_i}{2}$ (see (9)) expression (15) can be evaluated for every possible γ_q in an efficient manner.

This function contains the maximum information, and in contrast to the state of the art, the *a priori* knowledge and the soft decision information can fully be exploited. Thus the correlation between consecutive parameters is included in form of the *a priori*-probability. Equation (13) is evaluated as a product of the two factors depicted in Fig. 5.
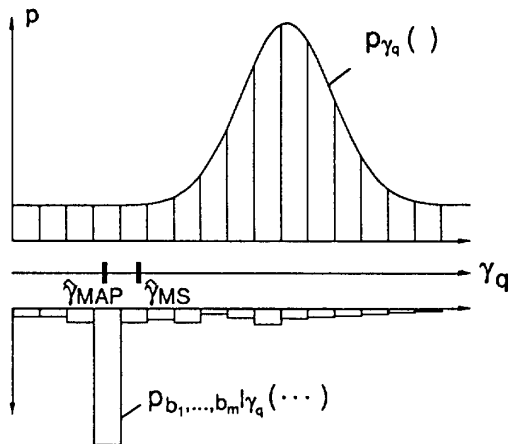
Figure 5: Components of *a posteriori*-probability.

Note:

Although the expressions above were derived for soft decision bits the same formalism can be applied if no channel state information is available. In that case the mean or estimated transition probabilities can be used for example.

5. BENEFIT

Equation (14) provides the basis to use optimum estimators. Standard candidates are the maximum *a posteriori* estimator (MAP) and the mean square estimator (MS). An example is indicated in Fig. 5 . It can be seen from Fig. 5 that if the channel becomes unreliable the transition probability $p_{b_1,...,b_m|\gamma_q}$ becomes flat and the *a priori*-probability p_{γ_q} gains influence. I. e. extrapolation takes place only if the reception is disturbed. An error detector is not required.

A very promising feature of the proposed method is that the estimator can be adapted to the perceptual influence of the specific parameter by applying a perceptual cost function.

6. APPLICATION

To study the new method we used a conventional CELP-algorithm with 8 kbps net bit rate and 3.4 kbps for error protection by punctured convolutional codes [6]. The whole transmission system was simulated with error patterns corresponding to $C/I = 10$, 7 and 4 dB. Using measured or approximated conditional densities for all existing coder parameters, the new approach was applied in the decoder with the use of soft decision values [7]. The speech decoder was realized on a DSP32C signal processor programmed in C.

In this first application the improvement of speech quality was significant. Especially under bad channel conditions (e. g. $C/I = 4$ dB) severe and annoying distortions were mostly avoided. It turned out, that for all parameters besides the pitch lag the MS-estimator worked better than the MAP–estimator proposed in [8].

7. SUMMARY AND CONCLUSIONS

A systematic approach for speech extrapolation in the decoder is given. With the extracted *a posteriori*-probability optimum estimators adapted to human perception can be applied and soft decision information can fully be exploited. It turns out, that in terms of perceptual performance the MS-estimator is superior to the MAP. The method was tested under realistic conditions using a 8 kbps CELP-codec. Significant improvement of subjective speech quality can be achieved.

REFERENCES

[1] ETSI/GSM, *Recommendations 06.11 Full rate lost speech frame substitution and muting*.

[2] P. Vary, K. Hellwig, D. Massaloux, J. Petit, C. Galand, and M. Rosso, "Speech codec for the european mobile radio system," in *IEEE Global Telecommunications Conference Dallas*, (Dallas, USA), November 27-30, 1989.

[3] H. yu Su and P. Mermelstein, "Improving the speech quality of cellular mobile systems under heavy fading," in *Proc. of ICASSP-92*, (San Francisco), pp. II 121-124, 1992.

[4] D. Sereno, "Frame substitution and adaptive post-filtering in speech coding," in *Proc. of Eurospeech*, (Genova), pp. 595-598, 1991.

[5] J. Hagenauer, "A viterbi algorithm with soft-decision outputs and its applications," in *Proc. of the GLOBECOM*, pp. 1680-1686, 1989.

[6] "High quality speech coding at medium to low bit rates," tech. rep., 1990. EUREKA PROJECT No. 151.

[7] H. C. Kniß. "Entwicklung eines Echtzeit-Interpolators", Diplomathesis at Institut für Nachrichtengeräte der RWTH Aachen, May 1991.

[8] K. Sayood and J. C. Borkenhagen, "Use of residual redundancy in the design of joint source/channel coders," *IEEE Trans. on Communications*, vol. 39, pp. 838-846, June 1991.