IMPLEMENTATION OF A ROBUST RELP SPEECH CODER

Harald Katterfeldt

AEG-Telefunken Research Institute
Ulm, Germany

Erich Behl

AEG-Telefunken Nachrichtentechnik GmbH.
Backnang, Germany

ABSTRACT

This paper presents a RELP coder based on baseband transmission and high-frequency regeneration in the frequency domain in two versions: 4.8 kb/s and 9.6 kb/s. The principle of both versions is illustrated in the detailed block diagram in Fig.1. Some details of a fixed-point implementation will be described. The 4.8 kb/s version has been implemented on a real-time processor. Since the algorithm does not distinguish between voiced and unvoiced speech, the number of calculations per frame is nearly constant. Thus a 4% time reserve is sufficient to insure real-time operation for the 4.8 kb/s version. The robustness of the RELP coder in the presence of background noise is explained. Except for narrow-band high-frequency background noise, the RELP coder transmits this noise as well as speech without any significant distortion. The influence of transmission errors is described. Simulations with a mobile radio channel model showed that the quality of the speech transmitted 'digitally' by RELP and a digital modem is definitely superior to that transmitted by simulated analog FM.

1. PRINCIPLE OF RELP CODER

The RELP coder (residual excited linear predictive coder) is presented in two versions: 4.8 kb/s (RELP48) and 9.6 kb/s (RELP96). The principle of both versions is explained in the first section with the detailed block diagram in Fig. 1.

- The input speech signal is sampled at 8 kHz by a 12-bit A/D converter.

- The digitized input signal x(n) is filtered by a tenth-order LPC inverse filter in lattice structure. The PARCOR coefficients of the 255-sample frames are determined by the autocorrelation method with an overlapping 384-point Hamming window.

- The LPC residual signal e(n) is converted to block-floating format with 13 bit mantissas. The block exponent is transmitted to the receiver. It can be interpreted as a coarse level value of the a 256-point FFT into the frequency domain.

- The 26 low-order Fourier coefficients E(k), k=1,...,26, which correspond to 32,...,832 Hz, form the 'baseband' which is transmitted to the receiver. The Fourier coefficients are first converted to polar coordinate representation. Then the magnitudes |E(k)| are encoded by an adaptive linear quantizer. The bit assignment and step size of the adaptive quantizer (linear gaussian type) is controlled by an estimated variance (k). The variance is estimated according to a homomorphic model proposed in /2/ and /3/.

- The cepstrum c(n) of the residuum is computed by the inverse DFT of the logarithmic magnitude values log |E(k)|. The pitch period is determined by searching the maximal cepstral value in the range from c(20) to c(128). The cepstral position p (which is the pitch period) and magnitude c(p) of the cepstral pitch peak are encoded and transmitted. This side information describes the periodic pitch structure of the residuum. The logarithmic mean value c(0) is also encoded and transmitted. It is used as the fine signal level. The estimated variance course log σ(k) is determined by the Fourier transform of the simplified cepstrum $\hat{c}$(n). This cepstrum consists of the decoded values $\hat{c}$(0) and $\hat{c}$(pitch) and zero for all other indices. Since only two values of the input signal $\hat{c}$(n) are not equal to zero, the Fourier transform is performed by staight-forward evaluation of DFT instead of FFT.

- The encoded magnitudes are decoded at the transmitter too to control the phase quantizer. The bits of the phase quantizer are assigned proportionally to the logarithm of the decoded magnitudes. Finally the phase values are encoded.

- In the 9.6 kb/s version the 'baseband' comprises the first 35 Fourier coefficients E(k) (32-1120 Hz). The untransmitted coefficients k > 35 are divided into 4 sub-bands, each consisting of 23 samples, corresponding to 736 Hz. Additional parameters for regeneration of these sub-bands are determined at the transmitter by cross correlation of the baseband with each sub-band /1/. 13 complex cross-correlation coefficients are computed per sub-band. The position of the maximal cross-correlation coefficient is detected and transmitted together with the angle of this coefficient. This side information represents the optimum-shift factors and phase corrections for regeneration of the upper bands.

CH1841-6/83/0000-1316 $1.00 © 1983 IEEE

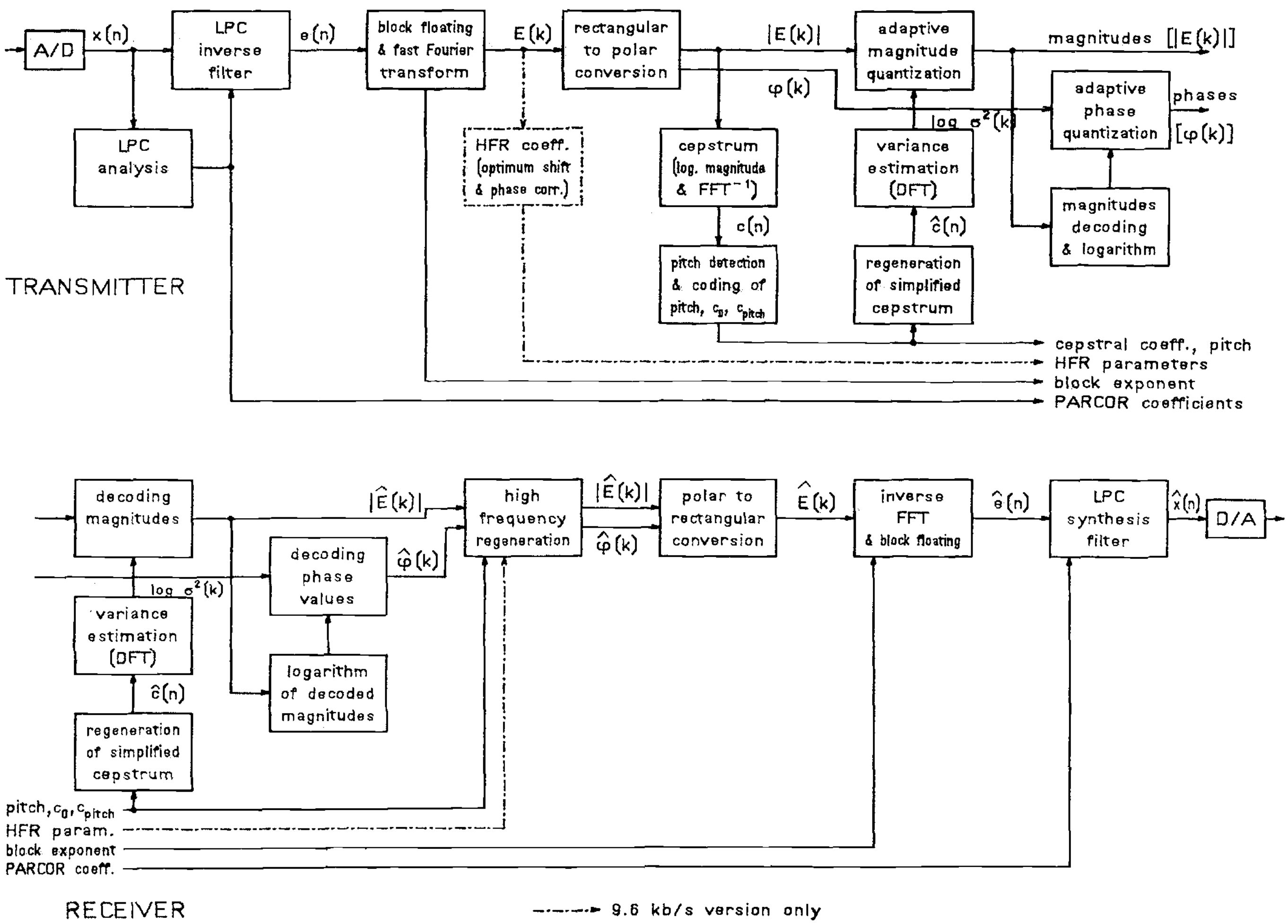
A/D
x(n)
LPC inverse filter
e(n)
block floating & fast Fourier transform
E(k)
rectangular to polar conversion
|E(k)|
adaptive magnitude quantization
magnitudes [|E(k)|]
φ(k)
adaptive phase quantization
phases [φ(k)]
log σ²(k)
LPC analysis
HFR coeff. (optimum shift & phase corr.)
cepstrum (log. magnitude & FFT⁻¹)
variance estimation (DFT)
magnitudes decoding & logarithm
c(n)
ĉ(n)
pitch detection & coding of pitch, c₀, c_pitch
regeneration of simplified cepstrum
TRANSMITTER
cepstral coeff., pitch
HFR parameters
block exponent
PARCOR coefficients
decoding magnitudes
|Ê(k)|
high frequency regeneration
|Ê(k)|
polar to rectangular conversion
Ê(k)
inverse FFT & block floating
ê(n)
LPC synthesis filter
x̂(n)
D/A
decoding phase values
φ̂(k)
φ̂(k)
log σ²(k)
variance estimation (DFT)
ĉ(n)
logarithm of decoded magnitudes
regeneration of simplified cepstrum
pitch, c₀, c_pitch
HFR param.
block exponent
PARCOR coeff.
RECEIVER
9.6 kb/s version only

Fig. 1 Block diagram of RELP coder

- At the receiver the base-band coefficients are first decoded, repeating the corresponding operations of the transmitter.

- The **untransmitted coefficients** are regenerated by shifting the base-band coefficients. In RELP48 the shift is controlled by the transmitted pitch value as explained in /1/. In RELP96, the shift is controlled by the transmitted optimum-shift parameters. A phase correction is included. No voiced/unvoiced decision is performed in either version because in the unvoiced case the pitch-adaptive HFR, which now is driven by a random pitch value, works more naturally and robustly than a random excitation signal would.

- The Fourier **coefficients** are converted back into cartesian coordinates.

- The **residual signal** is transformed back to the time domain.

- Finally, the **LPC synthesis filter** is excited by this residuum.

2. FIXED-POINT IMPLEMENTATION

These operations, most of which are single precision, are performed in 16-bit fixed-point arithmetic. Only the autocorrelation coefficients (during LPC analysis) and the cross correlation coefficients (for HFR parameters) are computed in 32 bit accuracy. The algorithm consists of many 'simple' operations such as multiplications, additions and shifts, and a small number of more complex operations such as divisions, logarithms, exponentations, sin, and cos. These more complex operations are performed as follows:

The **trigonometric functions**, which are required by the DFT and the polar-to-rectangular conversion, are performed by table lookup.

The **logarithm** to the base 2, which was chosen because of its ease of computation, is approximated as follows: The characteristic is the position of the leading binary '1' in the number. The mantissa is looked up in a logarithm table. The address for the table look up consists of the bits following the leading '1'.

Three exponentations per baseband sample are required for calculation of the step size of the linear magnitude quantizer and its inverse from the logarithmic variance values $\log \sigma^2(k)$. The exponentiations are performed by bitwise decomposition of the exponent:

$$2^x = 2^{x(n)} * 2^{x(n-1)} ... * 2^{x(i)} ... * 2^{x(0)}$$

where $x(i) = 0$ if bit i of the exponent is 0
and $x(i) = 2^{i-P}$ if bit i of the exponent is 1

P is the position of the binary point in the 16-bit word. The values of 2^{i-P} are stored in a table. The exponentation requires as many multiplications as there are bits of the exponent set to '1'.

The Fourier coefficients are converted with the CORDIC algorithm /4/ from rectangular to polar coordinates.

In addition to a FORTRAN simulation, the RELP48 coder is implemented on a 16-bit fixed-point real-time processor. The signal processor is based on the Am2901 slices and a fast 16*16 bit multiplier. The complete algorithm for transmitter and receiver requires 96% of the 32 ms time frame. The algorithm does not include any decisions or branches such as voiced/unvoiced, etc. So the required number of operations per frame is nearly constant. The number of cycles varies slightly, because the execution of operations such as block floating, CORDIC coordinate conversion, maximum search, and exponentation depends of the operands. Since the execution times vary only slightly, the reserve of 4% is sufficient to insure that the calculation unit does not exceed the available time frame.

3. ROBUSTNESS IN THE PRESENCE OF INPUT NOISE

This section is concerned with the behavior of the RELP coder if noise is added to the input speech. With respect to the base band, the RELP coder is an adaptive residual coder (ARC). Like other waveform coders, ARC relies very little on a speech production model. Therefore, it is not nearly as sensitive to input noise as vocoders. In the RELP48 coder, the base band is quantized relatively coarsely (3.6 bit per complex sample). The quantization noise results in a slight roughness of the synthetic speech. The roughness is speaker dependent. It is low if the speakers voice is suited to LPC-vocoder transmission. The roughness increases, if the voice does not lend itself to vocoder transmission or if proper analysis of speech-specific parameters, and hence the effective control of the adaptive quantizer, is affected by input noise. Likewise the block-boundary noise, which is typical for transform coding systems, increases. Input noise is transmitted relativly faithfully if at least a part of the spectrum of the noise is present in the base band. In most of these cases the increased quantization noise is not striking compared to the transmitted input noise.

Only high-frequent narrow-band noise is not transmitted particularly well. In the case of narrow-band noise, the LPC coefficients may adapt incompletely to the signal, which results in holes or peaks in the spektrum of the residuum. Fig. 2 shows two examples. An artifical rectangular signal (1.6 kHz) is added to the input speech. The spectra of the input signal with noise, the LPC filter, and the residuum before and after high-frequency regeneration are plotted in Fig. 2. The left example shows an 'over-adaption' of the LPC filter in which the spectrum of the residuum contains a zero. The zero is 'filled' by the HFR. Because of the pole in the LPC synthesis filter, this results in a loud whistling noise in the output signal of the coder. For this reason it is important to limit the maximal amplification of the LPC synthesis filter by limiting the magnitudes of the PARCOR coefficients during quantization.

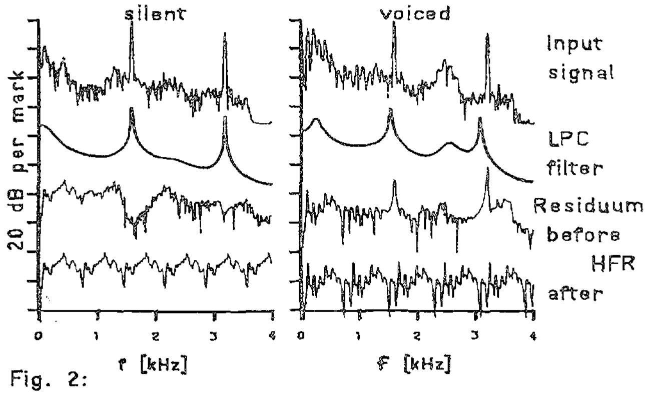

Fig. 2:
Influence of distortion by high-frequency rectangular signal

The left example occurs mainly in silent intervals. The right example shows a typical speech interval. The LPC filter is 'under-adapted', and the spectrum of the residuum contains peaks at fundamental frequency and first harmonic of the rectangular signal. These peaks are flattened by the HFR. After LPC synthesis the level of the noise is lower than in the input signal.

In this example of high-frequency narrow-band noise, the amplitude of the transmitted noise decreases below the input level during speech intervals, but increases up to and above the input level during silence intervals. This effect is unpleasant, because a constant whistling tone is more acceptable than a continuously wailing whistle. Except for this example we found that the RELP coder does not in itself significantly increase the deteriation of a speech signal caused by background noise.

4. ROBUSTNESS IN THE PRESENCE OF TRANSMISSION ERRORS

The sensitivity of main- and side-information coefficients to channel errors depends on their physical significance. In the subjective judgements given in the following, a statistically independent bit error rate of 10% is assumed.

- Errors in the pitch value cause an incorrect ripple structure in the estimated variance course $\tilde{\sigma}(k)$ for magnitude quantizer control and a pitch-asynchronous high-frequency regeneration. Because of the first effect, step sizes and bit assignments of transmitter and receiver do not agree. Therefore the base band values are decoded completely wrong. The excitation signal then looks like random noise. The synthetic speech sounds like whispering, but is still understandable. Compared to this effect, the deterioration due to the non-pitch-synchronous HFR is neglectible.

- If the received **block exponent** is erronuous, the signal level of the distorted frame is either too high or too low. Occasional errors resulting in lower levels are hardly perceptible, but the opposite case is very disturbing, because a loud crack occurs if a silent frame is synthesized with a high block exponent.

- Errors within the **PARCOR coefficients** result in a false synthesis filter which affects the spectrum of the output signal. Therefore, these errors impair the intellegibility more than errors in the other coefficients.

- Errors in the **cepstral coefficients** $\hat{c}(0)$ and $\hat{c}(\text{pitch})$ behave qualitatively like errors in block exponent and pitch value, but with far less effect on the speech quality.

- Errors in the **magnitudes of baseband coefficients** $|E(k)|$ result in increased roughness. In some cases they result in false bit assignment for the phase values too. Then the prevailing part of the phase values is decoded incorrectly. Then the above mentioned effect of whispering speech occurs.

- Errors in the **phase values** are hardly noticeable.

- Errors in the **HFR parameters** (RELP96) are neglectible.

Considering the combination of these effects, statistically independent error rates up to 10^{-3} are hardly perceptible. Error rates between 10^{-3} and 10^{-2} become noticeable, and above 10^{-2} intelligibility is affected.

To counteract this, an error protection code was implemented. The 11 most sensitive bits of a frame (most significant bits of pitch, block-float exponent and of the first tree PARCOR coefficients) are protected by a (15,11) Hamming code which is capable in correcting 1 error. Using this code a bit error rate of 10^{-2} in these 11 bits can be decreased to $2 * 10^{-3}$.

We tested the performance of the RELP coder simulating 'digital' and 'analog' speech transmission over a mobile-radio channel model. The channel model simulates Rayleigh-fading. The analog speech is 'transmitted' by idealized frequency modulation for a mobile radio channel. The digital speech signal (digitized by RELP) is 'transmitted' by band-limited minimum shift keying (BMSK) modulation similar to GMSK /5/. The average signal-to-noise ratio on the channel was adjusted to 14.7 dB., resulting in an average bit error rate of $8 * 10^{-3}$ at 9.6 kb/s and $4 * 10^{-3}$ at 4.8 kb/s. Comparing both systems, the FM distortions of 'analog' transmission turned out to be far more unpleasant than the influence of bit errors on the RELP-coded speech. Recorded examples of this will be presented.

5. SUMMARY

A 4.8 kb/s and 9.6 kb/s version of a fixed-point implementation of a DFT-based RELP coder have been presented. The 4.8 kb/s version runs on a real time processor with a time reserve of about 4%. The coder proved to be robust in the presence of background noise at the microphone input and transmission errors. Simulations with a mobile-radio channel model showed that in case of frequent channel disturbations speech transmitted 'digitally' by RELP in conjunction with a band-limited MSK modulator, has a significantly more acceptable quality than speech which was transmitted by 'analog' FM.

REFERENCES:

/1/ Harald Katterfeldt
A DFT-Based Residual Excited Linear Predictive Coder (RELP) for 4.8 and 9.6 kb/s
ICASSP'81 Atlanta, Georgia, conf. rep. pp. 824-827

/2/ Rainer Zelinski
Quellencodierung von Sprachsignalen mit Verfahren der adaptiven Transformationscodierung
9. Internationaler Kongress Mikroelektronik, Muenchen, 1980 Tagungsband S. 109-113

/3/ Richard V. Cox, Ronald E. Crochiere
Real-Time Simulation of Adaptive Transform Coding
IEEE Trans. on ASSP Vol. 2 April 81

/4/ J.E. Volder
The CORDIC trigonometric computing technique
IRE Trans. Electron. Comput. Vol EC-8, Sept.1959 pp. 330-334

/5/ Murota, K; Hirade, K.
GMSK Modulation for Digital Radio Telephony
IEEE Trans. Commun. Vol COM-29, No. 7,July 1981 pp. 1044-1050