FBG MODEL BASED LOW RATE CODING OF SPEECH

Kjeld Hermansen, Uwe Hartmann
Institute of Electronic Systems
DICOM group
Aalborg University
Fredrik Bajers Vej 7
DK-9220 Aalborg, Denmark
E-mail: kjh@kom.auc.dk, uh@kom.auc.dk

ABSTRACT
Speech coding algorithms are basic components in existing and future personal communication systems. Reducing the bit rate in transmitting speech is by way a matter of using slowly varying and numerical robust parameters to represent the speech. We are here focusing on parameters close to the physics of the speech production system for modelling the vocal tract. We have studied the dynamics (time variation) and numerical robustness of the new parameters (f formant frequency, b 3 dB bandwidth, g gain) and compared them with the traditional reflection coefficients. Listening tests indicate a possible reduction in bit rate of about two to three times for comparable sound quality. These fbg-parameters will be coded, transmitted, interpolated and used as input into the receiver model in order to reconstruct the speech signal.

INTRODUCTION
The role of speech coding has gained an increasing importance due to the advent of digital cellular telephones. With the low data rate speech coding algorithms, the speech quality tends to degrade drastically in high noise. Our speech-coding algorithm using the more noise robust fbg parameters allows a low data rate without affecting the speech quality and intelligibility drastically.

SIGNALPROCESSING
A speech signal can be modelled in different ways. One possibility is to use an autoregressive model, e.g. of order 10, and represent the speech signal by the LPC-coefficients or reflection coefficients. In our model the speech signal is represented by a set of fbg filters each corresponding to a resonance peak where f is the formant frequency, b the 3 dB bandwidth and g the gain of a resonant peak.

Figure 1 fbg tracks for the word 'zero'.

Figure 1 shows a visualization of the time development of four fbg filters of a pronunciation of the word ‘zero’. The four traces represent four formant frequencies f. The width of a trace is the 3 dB bandwidth b. The colour is the gain g of the peak.
Figure 2 LPC model and fbg filters.

Figure 2 shows the LPC-spectrum (upper curve) of a speech segment and some fbg filters of second order (lower curves). Usually three or four filters contain the essential information of the speech signal.

Whether we use reflection coefficients or fbg parameters, they both represent the LPC-model.

However, the fbg filters are closer to the physics of the speech producing organs than the reflection-coefficients. Due to the inertia of the speech producing system, the fbg-coefficients usually will have a slow time variation in contrast to the reflection-coefficients which can have a rapid time variation.

TIME QUANTIZATION

The speech signal is sampled with a sampling frequency of 8 kHz. A framelength of 30 msec corresponding to 240 samples is used for calculating the model parameters of the speech segment and updated with a frame shift of 1.25 msec corresponding to 10 samples.

Figure 3 First formant and second reflection coefficient for the word 'zero'.

The parameters of the speech are considered as time varying signals.

Figure 3 shows the first formant frequency and the second reflection coefficient for a pronunciation of the word ‘zero’. The formant frequency is slowly time varying whereas the reflection-coefficient varies substantially.

Figure 4 Spectra of 1st formant and 2reflection coefficient for 'zero'.

The upper curve in figure 4 shows the formant spectrum and the lower curve the spectrum of the reflection coefficient. The curves indicate a decrease in bandwidth of approximately a factor two using fbg-coefficients in stead of reflection coefficients.
This section describes the data reduction using fbg parameters in stead of reflection coefficients. To have comparable sound quality, we use the three dominant formant frequencies for fbg representation and a tenth order all pole filter for reflection coefficient representation. Coding the residual signal is the same for the two representations so this is not included. Every frame has consequently 9 and 11 coefficients respectively. Due to the lower bandwidth (factor 2) for the fbg parameters, the frame sampling frequency can be 30 Hz and 60 Hz respectively. The number of coefficients per second are 9*30=270 and 11*60=660 respectively. All fbg parameters are positive so unsigned representation can be used. In average – f, b and g parameters – 5 bit give no audible quantization effects. For the reflection coefficients which have a dynamic range in the interval from –1 to 1, the corresponding average word length is 7 bit.

CONCLUSION
This paper has presented a new method for speech coding using fbg-parameters giving a reduced data rate due to the slow variation and the numerical robustness of the parameters. The method is more time consuming than standard methods but still suitable for real time implementation.

REFERENCES


(3) Takashi Ikeda, Akira Watanabe, Yuichi Ueda, Kouji Tasaki: Characteristics of a Hearing Aid for Telephonic Speech by Single Resonant Analysis. 15th International
Congress on Acoustics. Trondheim, Norway

(4) Hartmann U., Hermansen K.: Model based
Spectral Subtraction in Speech with low SNR.
NORSIG2000, Kolmården, Sweden June 13 –
June 15 2000.