Amplitude - temporal method of speech coding

Victor Ababii, Viorica Sudacevschi
Technical University of Moldova
E-mail: ababii@mail.utm.md

ABSTRACT

A method of speech coding and decoding is proposed. The speech coding algorithm is based on first derivate calculation of input speech signal, identification of critical points and input signal amplitude in these points, time period measurement between critical points. The result of codification represents a sequence of amplitudes and time periods. The decoding algorithm utilizes values of COS or SIN functions for reconstruction of the input speech. The codec structure that consists from encoder and decoder units is proposed.

Keyword: amplitude, coding, identification, modulation, signal, speech, temporal

1. INTRODUCTION

Narrowband speech codecs are used to give an efficient digital representation of telephone bandwidth speech. Often the speech is band limited to between 200 and 3400 Hz, and is sampled at 8 kHz. An ideal speech codec will represent this speech with as few bits as possible, while producing reconstructed speech which sounds identical, or *almost* identical, to the encoded speech. Of course in practice there is always a trade-off between the bit rate of the codec and the quality of its reconstructed speech [1].

Speech codecs are often broadly divided into three classes - waveform codecs, source codecs and hybrid codecs. Typically waveform codecs are used at high bit rates, and give very good quality speech. Source codecs operate at very low bit rates, but tend to produce speech which sounds synthetic. Hybrid codecs use techniques from both source and waveform coding, and give good quality speech at intermediate bit rates.

The simplest form of waveform coding is Pulse Code Modulation (PCM), which merely involves sampling and quantizing the input waveform (standard G 711). Narrow-band speech is typically band-limited to 4 kHz and sampled at 8 kHz. If linear quantization is used then to give good quality speech around twelve bits per sample are needed, giving a bit rate of 96 kbits/s. This bit rate can be reduced by using non-uniform quantization of the samples [3].

A commonly used technique in speech coding is to attempt to predict the value of the next sample from the previous samples [2]. If the predictions are effective then the error signal between the predicted samples and the actual speech samples will have a lower variance than the original speech samples. This is the basis of Differential Pulse Code Modulation (DPCM) schemes - they quantize the *difference* between the original and predicted signals.

Adaptive Differential Pulse Code Modulation (ADPCM) codecs are waveform codecs which instead of quantizing the speech signal directly, like PCM codecs, quantize the difference between the speech signal and a prediction that has been made of the speech signal. In the mid 1980s the CCITT standardized a 32 kbits/s ADPCM, known as G721, which gave reconstructed speech almost as good as the 64 kbits/s PCM codecs. Later in recommendations G726 and G727 codecs operating at 40, 32, 24 and 16 kbits/s were standardized.

Waveform coders are capable of providing good quality speech at bit rates down to about 16 kbits/s, but are of limited use at rates below this. Source codecs on the other hand can provide intelligible speech at 2.4 kbits/s and below, but cannot provide natural sounding speech at any bit rate. Although other forms of hybrid codecs exist, the most successful and commonly used are time domain Analysis-by-Synthesis (AbS) codecs. AbS codecs were first introduced in 1982 by Atal and Remde with what was to become known as the Multi-Pulse Excited (MPE) codec. Later the Regular-Pulse Excited (RPE) and the Code-Excited Linear Predictive (CELP) codecs were introduced.

Information Technologies 2004, edited by Andrei M. Andriesh, Veacheslav L. Perju,
Proceedings of SPIE Vol. 5822 (SPIE, Bellingham, WA, 2005)
0277-786X/05/$15 · doi: 10.1117/12.612046