MELP Vocoder Algorithm

The New 2400 bps Federal Standard Speech Coder

Overview

The MELP (Mixed-Excitation Linear Predictive) Vocoder is the new 2400 bps Federal Standard speech coder. It was selected by the United States Department of Defense Digital Voice Processing Consortium (DDVPC) after a multi-year extensive testing program. The selection test concentrated on four areas: intelligibility, voice quality, talker recognizability, and communicability. The selection criteria also included hardware parameters such as processing power, memory usage, and delay. MELP was selected as the best of the seven candidates and even beat the FS1016 4800 bps vocoder, a vocoder with twice the bit-rate.

MELP is robust in difficult background noise environments such as those frequently encountered in commercial and military communication systems. It is very efficient in its computational requirements. This translates into relatively low power consumption, an important consideration for portable systems.

The MELP Vocoder was developed by a team from Texas Instruments Corporate Research in Dallas and Atlanta Signal Processors. The MELP Vocoder is based on technology developed at the Center for Signal and Image Processing at the Georgia Institute of Technology in Atlanta.

General Description

Traditional pitched-excited LPC vocoders use either a periodic pulse train or white noise as the excitation for an all-pole synthesis filter. These vocoders produce intelligible speech at very low bit rates, but they sometimes sound mechanical or buzzy and are prone to annoying thumps and tonal noises. These problems arise from the inability of a simple pulse train to reproduce all kinds of voiced speech. The MELP Vocoder uses a mixed-excitation model that can produce more natural sounding speech because it can represent a richer ensemble of possible speech characteristics. The MELP Vocoder is also robust in difficult background noise environments such as those frequently encountered in commercial and military communication systems.

The MELP Vocoder Algorithm

The MELP Vocoder is based on the traditional LPC parametric model, but also includes four additional features. These are mixed-excitation, aperiodic pulses, pulse dispersion, and adaptive spectral enhancement.

The mixed-excitation is implemented using a multi-band mixing model. This model can simulate frequency dependent voicing strength using a novel adaptive filtering structure based on a fixed filterbank. The primary effect of this multi-band mixed-excitation is to reduce the buzz usually associated with LPC vocoders, especially in broadband acoustic noise.

When the input speech is voiced, the MELP vocoder can synthesize speech using either periodic or aperiodic pulses. Aperiodic pulses are most often used during transition regions between voiced and unvoiced segments of the speech signal. This feature allows the synthesizer to reproduce erratic glottal pulses without introducing tonal noises.

The pulse dispersion is implemented using fixed pulse dispersion filter based on a spectrally flattened triangle pulse. This filter has the effect of spreading the excitation energy with a pitch period. This, in turn, reduces the harsh quality of the synthetic speech.

The adaptive spectral enhancement filter is based on the poles of the LPC vocal tract filter and is used to enhance the formant structure in the synthetic speech. This filter improves the match between synthetic and natural bandpass waveforms, and introduces a more natural quality to the speech output.

The first ten Fourier magnitudes are obtained by picking peaks in the FFT of the residual signal. The information embodied in these coefficients improves the accuracy of the speech production model at the perceptually important lower frequencies. This increases the quality of the coded speech, particularly for males and in the presence of background noise.
Availability

The MELP algorithm is now available in ANSI C. MELP Object code optimized for the Texas Instruments TMS320C3x is also available. For information on availability on other processors, contact an Atlanta Signal Processors representative.

Specifications

**Hardware Configuration**
- Microprocessor: Texas Instruments TMS320C3x Digital Signal Processor, operating at 50 MHz or above
- Memory requirements: 15,600 words program, 4,900 words data, plus 1,130 words on-chip RAM
- 50 MHz 'C30 utilization, full duplex: 82% at 2400 bps
- 50 MHz 'C30 utilization, encode: 59% at 2400 bps
- 50 MHz 'C30 utilization, decode: 23% at 2400 bps

**Algorithm**
- Source language: C, with extensive sections in hand-optimized TMS320C3x assembly language
- Data rate: 2400 bps
- Sampling rate: 8 kHz
- Signal input: 16-bit linear
- Bit stream format: For each 22.5 ms frame of input speech, the following 54 bits are placed into the bit-stream (in this order)
  - **Description**
  - **Number of bits**
    - Pitch index
    - Jitter flag
    - Bandpass voicing decision
    - Gain for second half of frame
    - Gain for first half of frame
    - LSP frequencies (10 line spectrum pairs)
    - Fourier magnitudes (10 harmonies)
    - Sync bit
    - Total
    - 7
    - 1
    - 4x1
    - 5
    - 3
    - 25
    - 8
    - 1
    - 54

The speech is broken down into frames of 180 samples (44,444 frames per second).

The bit rate is given by: 54 \* 44,444 = 2400 bps