VoiceAge Corporation
750 Chemin Lucerne, Suite 250
Ville Mont-Royal (Quebec)
H3R 2H6 Canada
(514) 737-4940 Fax (514) 908-2037
www.voiceage.com

Open G.729 Initiative

Technical Documentation

Version 1.0 – Revision 2004-07-15

Table of Contents

PACKAGE CONTENTS 3

INPUT/OUTPUT FORMAT 4

INPUT SPEECH LEVEL FOR OPTIMUM PERFORMANCE...5

HANDLING TRANSMISSION BIT ERRORS AND FRAME LOSSES..6

ABOUT THE ENCODER/DECODER SAMPLE PROGRAMS 6

G729A API FUNCTIONS...7

FAQS 9

Open G.729 Initiative

Technical Documentation

VoiceAge G.729 is an 8 Kbps coder that encodes/decodes speech signals using the Conjugate-Structure Algebraic-Code-Excited-Linear-Prediction (CS-ACELP) algorithm. VoiceAge G.729 is a reduced complexity version of G.729 and is bit stream interoperable with the full version.

The coder operates on speech frames of 10 ms, corresponding to 80 samples at a sampling rate of 8000 samples/sec. In addition to the 10 ms speech frame duration, there is also a look-ahead delay of 5 ms, resulting in a total initial algorithmic delay of 15ms.

VoiceAge G729 codec specifications	
Bit rate (kbps)	8
Speech sampling rate (Hz)	8000
Frame duration (ms)	10
Look-ahead delay (ms)	5

PACKAGE CONTENTS

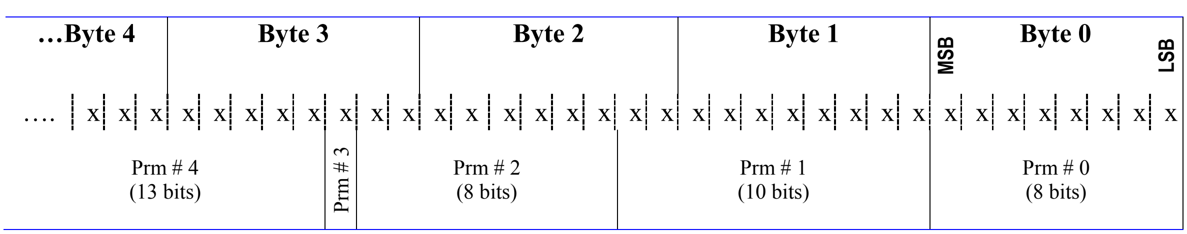
va_td_g729.pdf	This document.
va_g729a.lib	Win32 statically linkable library of G729 floating-point object code for Pentium and compatible processors.
va_g729a.h	API prototypes and constants declarations required by the sample programs.
va_g729a_encoder.c	Encoder sample application demonstrating encoder API calls to the codec for encoding a speech file.
va_g729a_decoder.c	Decoder sample application demonstrating decoder API calls to the codec for decoding a speech file.
va_g729a_encoder.exe	Encoder sample program executable for the Win32 platform.
va_g729a_decoder.exe	Decoder sample program executable for the Win32 platform.

INPUT/OUTPUT FORMAT

The encoder requires raw 16-bit mono PCM speech data sampled at 8000 Hz as input, i.e., without any header information. For every speech frame, consisting of 80*16 bit (160 bytes) samples, the encoder produces a 10*8 bit (10 bytes) packed parameter list that can be transmitted to the decoder. At the decoder, these parameters are used to reconstruct the speech data frame in the same format as the encoder input.

The parameter list produced by encoding a frame of audio data is packed into 80 bits (10 bytes). The bits allocated for each parameter are converted to a stream of bytes that are transmitted from the rightmost to leftmost byte as described in the following tables:

Parameter	Description	Bits Used
prm #0	1st codebook	8 bits
prm #1	2nd codebook	10 bits
		1st subframe
prm #2	pitch period	8 bits
prm #3	parity check on 1st period	1 bit
prm #4	codebook index1 (positions)	13 bits
prm #5	codebook index2 (signs)	4 bits
prm #6	pitch and codebook gains	7 bits
		2nd subframe
prm #7	pitch period (relative)	5 bits
prm #8	codebook index1 (positions)	13 bits
prm #9	codebook index2 (signs)	4 bits
prm #10	pitch and codebook gains	7 bits



INPUT SPEECH LEVEL FOR OPTIMUM PERFORMANCE

A speech sample in the digital domain is represented by a 16-bit 2's complement value. Thus, the overload point is 32767. How this overload point refers to the analog world depends on the conversion method between the analog and digital domain. If a 16-bit A/D is used, the samples are already represented by 16-bit integers. If this is not the case, then the A/D output should be properly shifted to fit in the 16-bit representation. For example, if a 14-bit A/D is used, the data should be shifted to the left by 2 before reading the data into the input speech buffer.

The power of a signal x(n) with a length of N samples is defined by:

$$P = 1/N\sum_{n=0}^{N-1} x^2(n)$$

The power level in decibels is defined relative to a reference power level P0 = 32768.0 as:

$$L = 10\log_{10}(P/P_0)$$

Note that reference level P0 (L=0 dBov) corresponds to a DC signal with amplitude equal to the overload point 32767. Note also that the level of a sine wave with a peak value of 32767 is L=-3.01 dB. The nominal input level of the CODEC is Pn = -25 dB. In general, this corresponds to speech data with an absolute maximum of approximately 16000. The CODEC performs best in the input level range Pn ± 5 dB. At higher input levels (say -15 dB) the input signal is occasionally saturated at 32767, and this results in some degradation in CODEC performance. It is advised that the input level of the A/D should be properly tuned to insure that the power level at the CODEC input is in the range Pn ±5 dB.

CODEC COMPLEXITY

This floating-point implementation of G.729 Annex A is based on the fixed-point G.729 Annex A reduced complexity speech codec. The complexity is represented as percentage of CPU usage, and is as follows when tested on an Intel 500 MHz Celeron-MMX:

Encoder	6% CPU time
Decoder	2% CPU time

HANDLING TRANSMISSION BIT ERRORS AND FRAME LOSSES

If a speech frame is lost (as in packetized speech transmission) or a frame is badly corrupted (as in mobile radio transmission), the decoder can perform a frame substitution using the parameters received in the last frame.

The parameter **bfi** is used at the decoder to indicate frame erasures. This parameter should always be set to 0 under normal conditions. To signal a frame erasure, **bfi** can be set to 1 by an external error detector, thus signaling to the decoder a frame loss or low confidence in the received frame. In that case, the decoder will apply frame substitution.

It is also possible to protect the **lsf**, **pitch**, and **gains** parameters independently, so that an error in one of these parameters does not cause a frame erasure (global error).

The following table describes how to signal an error with bfi:

Bad Frame Indicator	Error Type	Affected Parameters
Bfi = 0	No error	
Bfi = 1 (bit0)	Erasure (global error)	(all prm)
Bfi = 2 (bit1)	Error in lsf parameters	(prm #0 & #1)
Bfi = 4 (bit2)	Error in pitch parameters	(prm #2, #3 & #7)
Bfi = 8 (bit3)	Error in gains parameters	(prm #6 & #10)

ABOUT THE ENCODER/DECODER SAMPLE PROGRAMS

The sample programs va_g729a_encoder.c and va_g729a_decoder.c are used to simulate the encoder and decoder, and demonstrate how to initialize and call the encoding and decoding process.

The encoder and decoder are run as follows (where **infile** and **outfile** are raw 16 bit PCM files sampled at 8 kHz):

```
VA_G729A_ENCODER infile bitstream
VA_G729A_DECODER bitstream outfile
```

To build the speech encoder (or decoder) sample programs, compile the file va_g729a_encoder.c (or va_g729a_ecoder.c). Link this object file to the codec library va_g729a.lib.

G729A API FUNCTIONS

va_g729a_init_encoder

Description	Initializes the static memory needed by the encoding process. This function must be called prior to opening or re-opening a channel.
Syntax	#include "va_g729a.h" void va_g729a_init_encoder(void);
Arguments	none
Returned value	none

va_g729a_encoder

Description	Encodes an 80 words speech frame into a 10 bytes packed bit stream.
Syntax	#include "va_g729a.h" void va_g729a_encoder(short *speech, unsigned char *bitstream);
Arguments	speech: Input speech buffer containing one frame of 16-bit PCM speech data. bitstream: Output bit stream buffer containing packed bit stream.
Returned value	none

va_729a_init_decoder

Description	Initializes the static memory needed by the decoding process. This function must be called prior to opening or re-opening a channel.
Syntax	#include "va_g729a.h" void va_g729a_init_decoder();
Arguments	speech: Input speech buffer containing one frame of 16-bit PCM speech data. bitstream: Output bit stream buffer containing packed bit stream.
Returned value	none

va_g729_ decoder

Description	Decodes a 10 bytes packed bit stream into an 80 words speech frame. The decoder can handle frame erasures through the bfi parameter.	
Syntax	#include “va_g729a.h” void va_g729a_decoder(unsigned char *bitstream, short *synth, int bfi);	
Arguments	bitstream:	Input buffer containing packed bit-stream.
	synth:	Output buffer containing one frame of decoded 16 bits PCM speech data.
	bfi:	Used to indicate a frame loss to the decoder. Set bfi=1 to have decoder perform a frame substitution using the previous frame.
Returned value	none	

FAQs

Here are some frequently asked questions about the floating-point implementation of G729 Annex A.

Q — Is the floating-point implementation of G.729 Annex A interoperable with the fixed-point version?

A — The floating-point version of G.729a is bit-stream identical to fixed-point version of G.729a. Therefore, a floating-point decoder can be used with a fixed-point encoder, and vice-versa.

Q — What type of speech input format is required?

A — Raw 16-bit mono PCM sampled at 8000Hz. Do not use .WAV files. They contain a header that will produce distortion at the start of a decoded audio sample because the encoder interprets the header as speech data.

Q — When I compress and decompress a file, why is the output file not the same size as the input file?

A — The G.729a codec operates on speech frames of 80 samples/frame, 16 bits (integers), i.e., 160 bytes/frame. The output file is always a multiple of 160 bytes, the size of one frame.

Q — How can I convert my .WAV files to raw 16 bit mono PCM sampled at 8000 Hz?

A — Use an audio editing tool such as SoX - Sound eXchange. See home.sprynet.com/~cbagwell/sox.html for more information

Q — Can I get link on platforms other than Pentium or compatible?

A — The object code provided in this package is Microsoft Win32 compatible and is compiled for the Pentium family of processors.

Q — Is the G.729a codec able to handle multiple channels?

A — No, only a single instance of the coder (or decoder) can be used at a time.

Q — The complexity of the codec seems higher than the values specified in the documentation. Why?

A — The values are based on a Celeron-MMX 500MHz.