Application Note
Using the Voice Activity Detector (VAD) in the TLV320ADCx120 and PCMx120-Q1

TEXAS INSTRUMENTS

Muthuraj Pillai, Diljith Thodi

Low Power Audio

ABSTRACT

The TLV320ADCx120/PCMx120-Q1 family of devices (TLV320ADC3120 /PCM3120-Q1 and TLV320ADC5120/ PCM5120-Q1 and TLV320ADC6120/PCM6120-Q1) are dual-channel, high-performance, analog-to-digital converters (ADC) for audio applications. This family of devices has an extensive set of features that include the following:

- Programmable channel gain
- Digital volume control
- Programmable microphone bias voltage
- Phase-locked loop (PLL)
- Programmable high pass filter (HPF)
- Automatic gain control (AGC)
- Dynamic range enhancer (DRE) support in the TLV320ADC5120 /PCM5120-Q1 and TLV320ADC6120/ PCM6120-Q1
- Linear phase or low-latency filter modes for sample-rates up to 768 kHz
- Voice activity detector (VAD)

This application note describes how to configure the Voice Activity Detector (VAD) feature in TLV320ADCx120/ PCMx120-Q1 devices.

Table of Contents

List of Figures

List of Tables

Trademarks

All trademarks are the property of their respective owners.

1 Introduction

Voice Activity Detector (VAD) is a voice-triggered system wake-up mechanism. VAD enables the rest of the system to be in sleep mode in the absence of voice activity, thereby consuming very low power. VAD-based systems generate an interrupt on detection of voice activity. Figure 1-1 shows how the VAD responds to voice activity.

Figure 1-1. VAD Example

VAD is supported on all analog-to-digital converter (ADC) channels of the TLV320ADCx120/PCMx120-Q1 device family, that is, on analog and digital microphone channels. The digital microphone channel is preferred for low-power applications. This application note describes the operation of the VAD, the tunable parameters, and the device configurations required to support VAD.

2 Voice Activity Detector

The VAD block monitors the signal from a microphone channel for voice-like patterns and on detection of a matching pattern that triggers an interrupt. The VAD monitors for both an onset of voice-activity as well as the end of voice-activity. Both events can be mapped to interrupts.

The ADCx120 device also has the capability to automatically power-on and power-off based on the VAD interrupts. As an example, the ADCx120 system can be set up to monitor VAD activity on a digital microphone channel and then power on the analog microphone channels based on the VAD trigger.

There are two modes of VAD operation:

1. Auto Mode: The VAD triggers automatically to power on and power off the system.
2. User Mode: The VAD interrupts have to be monitored by the host and the system has to be powered on or off through I2C commands.

Note that in both Auto and User modes, the device also generates an interrupt on the GPIO or GPO pin which can be sent to an external DSP or SOC.

The salient features of VAD are as follows:

*   **No need for external clocks** - The system reuses an internal oscillator tuned to generate the ADC and DSP clocks of the required frequency.
*   **Automatic switching between VAD mode and *Record* mode** - The system switches from VAD mode to Record mode upon voice activity and switches back to VAD mode upon no voice activity in Auto VAD mode.
*   **Reconfigurable decision tree from RAM** - This feature allows the default voice-activity decision tree to be updated with a different decision tree that has been trained for other acoustic event detections.

The VAD algorithm uses a decision tree classification-based algorithm for voice activity detection. The decision-tree parameters can be updated through coefficient writes so the VAD block can be reconfigured for other applications that need to use a decision-tree for detection. A 16-band non-uniformly-spaced IIR filter-bank is used for feature extraction. Feature selection parameters and the decision tree can be computed offline and updated through coefficient-memory writes. Figure 2-1 shows the signal processing chain for VAD.
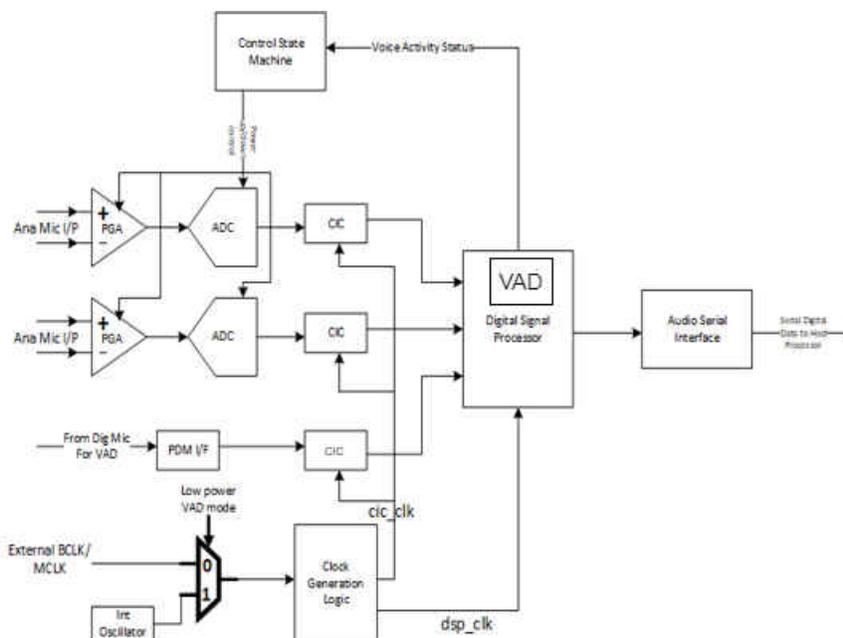

Figure 2-1. VAD Block Diagram

2.1 VAD Configurations

Table 2-1 shows the different modes in which VAD can be operated.

Table 2-1. List of VAD Configuration

VAD Configuration	Function, Description
User, Auto	Auto: ADC power up and power down happens automatically on voice detection and no voice, respectively. User: On detection of voice, interrupt is generated, user can initiate ADC power up and power down.
VAD channel	Decides which channel is to be monitored for VAD activity.
VAD clock	VAD needs to be run on internal clock or external clock.
VAD ON during recording	This function decides if voice detection needs to be active when recording is in progress.
SDOUT interrupt	SDOUT pin is enabled to support interrupt output when channel data is not being recorded.

User, Auto: VAD can be programmed by the user to be either in auto mode or user mode. There are 4 possible modes in which VAD can be programmed by the user.

0d = User initiated ADC power-up and ADC power-down: The user can initiate ADC power-up and ADC power-down based on the interrupt generated by the VAD algorithm.

1d = VAD interrupt based on ADC power up and ADC power down: This is the Auto mode, in which ADC is turned ON or OFF automatically based on the interrupt generated by the VAD algorithm.

As Table 2-2 shows, VAD mode selection is done using the VAD_mode[1:0] bit of VAD_CFG1 register (page = 0x01, address = 0x1E).

Table 2-2. VAD Mode Selection Using VAD_CFG1 Register

Bit	Field	Type	Reset	Description
7-6	VAD_MODE[1:0]	R/W	00b	Auto ADC power up and power down configuration selection. 0d = User initiated ADC power-up and ADC power down 1d = VAD interrupt based ADC power up and ADC power down 2d = VAD interrupt based ADC power up but user initiated ADC power down 3d = User initiated ADC power-up but VAD interrupt based ADC power down

VAD channel: This parameter decides which channel is to be monitored for VAD activity. Only one of the channels can be monitored for VAD activity at a time

As Table 2-3 shows, VAD channel selection is done using the VAD_CH_SEL[1:0] bit of VAD_CFG1 register (page = 0x01, address = 0x1E).

Table 2-3. VAD Channel Selection Using VAD_CFG1 Register

Bit	Field	Type	Reset	Description
5-4	VAD_CH_SEL[1:0]	R/W	10b	VAD channel select. 0d = Channel 1 is monitored for VAD activity 1d = Channel 2 is monitored for VAD activity 2d = Channel 3 is monitored for VAD activity 3d = Channel 4 is monitored for VAD activity

VAD clock: VAD can be run on either the internal oscillator clock or the external clock provided by the user. This external clock can be given on either the BCLK pin or the MCLK pin.

As Table 2-4 shows, VAD clock selection is done using the VAD_CLK_CFG[1:0] bit of VAD_CFG1 register (page = 0x01, address = 0x1E). If the user selects either 01b or 10b, then the frequency of external clock is selected using VAD_EXT_CLK_CFG[1:0] bit of VAD_CFG1 register (page = 0x01, address = 0x1E) as shown in Table 2-5.

Table 2-4. VAD Clock Selection Using VAD_CFG1 Register

Bit	Field	Type	Reset	Description
3-2	VAD_CLK_CFG[1:0]	R/W	00b	Clock select for VAD 0d = VAD processing using internal oscillator clock 1d = VAD processing using external clock on BCLK input 2d = VAD processing using external clock on MCLK input 3d = Custom clock configuration based on MST_CFG, CLK_SRC and CLKGEN_CFG registers in page 0

Table 2-5. VAD Clock Frequency Selection Using VAD_CFG1 Register

Bit	Field	Type	Reset	Description
1-0	VAD_EXT_CLK_CFG[1:0]	R/W	00b	Clock configuration using external clock for VAD. 0d = External clock is 3.072 MHz 1d = External clock is 6.144 MHz 2d = External clock is 12.288 MHz 3d = External clock is 18.432 MHz

VAD ON during recording: This parameter decides if voice activity needs to be detected when ADC is recording is on going or not. If this bit is enabled, then the VAD algorithm continues running when ADC recording is in progress to detect any voice activity.

As Table 2-6 shows, VAD ON during recording selection is done using the VAD_PD_DET_EN bit of VAD_CFG2 register (page = 0x01, address = 0x1F).

Table 2-6. VAD ON During Recording Selection Using VAD_CFG2 Register

Bit	Field	Type	Reset	Description
3	VAD_PD_DET_EN	R/W	1b	Enable ASI output data during VAD activity. 0d = VAD processing is not enabled during ADC recording 1d = VAD processing is enabled during ADC recording and VAD interrupts are generated as configured

SDOUT as interrupt: When ADC recording is not in progress, the SDOUT pin can be used for VAD interrupt. Setting this bit enables SDOUT to be used as the VAD interrupt pin.

As Table 2-7 shows, SDOUT as interrupt selection is done using the SDOUT_INT_CFG bit of VAD_CFG2 register (page = 0x01, address = 0x1F).

Table 2-7. SDOUT as Interrupt Selection Using VAD_CFG2 Register

Bit	Field	Type	Reset	Description
6	SDOUT_INT_CFG	R/W	0b	SDOUT interrupt configuration. 0d = SDOUT pin is not enabled for interrupt function 1d = SDOUT pin is enabled to support interrupt output when channel data in not being recorded

2.2 VAD Parameters

Table 2-8 shows the parameters of the VAD algorithm. These parameters reside in the 32-bit wide coefficient memory (Book 0, Page 1, Page 7, page 8. and Page 9) of the device.

Table 2-8. List of VAD Parameters

VAD Parameter	Function, Description
Initial learning period (ILP)	This is the amount of time the VAD algorithm takes to adjust itself to the background noise environment, from the time instant VAD is turned on.
Hold over counter (HOC)	On detecting voice activity, the VAD algorithm generates an interrupt, if the interrupt is programmed to be active high, then the interrupt goes high (logic 1) on detecting voice and goes low (logic 0) when there is no voice. Before going low, the amount of time interrupt stays high after the voice activity ceases to exist is determined by the HOC count.
Wakeup wait (WW)	If VAD is programmed to be in Auto mode, on detecting voice, it will automatically turn on the ADC and start recording, simultaneously also checking for voice activity. Wakeup wait is the amount of time for which VAD is suspended after going to recording mode and thereafter resumed.
Threshold (TH)	Threshold controls the decision boundary of the nodes of the decision tree. A higher value will increase the node thresholds of all the nodes of the decision tree, thus reducing the likelihood of false positives. Similarly, a lower value for the threshold parameter decreases the node thresholds, which reduces the likelihood of false negatives.

Initial learning period: ILP is the amount of time the VAD algorithm takes to adjust itself to the background noise environment, from the time instant VAD is turned on. Equation 1 shows the computation of the VAD_ILP parameter.

$$Initial\ learning\ period\ \Big(s\Big) = \frac{ILP_{10}}{\left(256 \times 8000\right)} \tag{1}$$

where

- ILP_{10} is the ILP register value in decimal form

The default value (0x001F4000) corresponds to 1 s. Table 2-9 shows the registers that control the VAD_ILP parameter.

Table 2-9. Programmable Coefficient Registers for Initial Learning Period

Coefficient	Page	Register	Reset Value	Description
VAD_ILP	0x07	0x7C	0x00	ILP Byte[31:24]
	0x07	0x7D	0x1F	ILP Byte[23:16]
	0x07	0x7E	0x40	ILP Byte[15:8]
	0x07	0x7F	0x00	ILP Byte[7:0]

Hold over counter: On detecting voice activity, the VAD algorithm generates an interrupt, if the interrupt is programmed to be active high, then the interrupt goes high (logic 1) on detecting voice and goes low (logic 0) when there is no voice. Before going low the amount of time interrupt stays high after the voice activity ceases to exist is determined by the hold over counter count. Equation 2 shows the computation of the VAD_HOC parameter.

$$Hold\ over\ counter\ \Big(s\Big) = \frac{HOC_{10}}{\left(256 \times 8000\right)} \tag{2}$$

where

- HOC_{10} is the HOC register value in decimal form

The default value (0x00032000) corresponds to 100 ms. Table 2-10 shows the registers that control the VAD_HOC parameter.

Table 2-10. Programmable Coefficient Registers for Hold Over Counter

Coefficient	Page	Register	Reset Value	Description
VAD_HOC	0x08	0x0C	0x00	HOC Byte[31:24]
	0x08	0x0D	0x03	HOC Byte[23:16]
	0x08	0x0E	0x20	HOC Byte[15:8]
	0x08	0x0F	0x00	HOC Byte[7:0]

Wakeup wait: If VAD is programmed to be in auto mode, on detecting voice, it will automatically turn on the ADC and start recording, simultaneously also checking for voice activity. Wakeup wait is the amount of time for which VAD is suspended after going to recording mode and thereafter resumed. Equation 3 shows the computation of the VAD_WW parameter.

$$Wakeup\,wait\,(s) = \frac{WW_{10}}{(256 \times 8000)} \tag{3}$$

where

- WW_{10} is the Wakeup wait register value in decimal

The default value (0x01388000) corresponds to 10 s. Table 2-11 shows the registers that control the VAD_WW parameter.

Table 2-11. Programmable Coefficient Registers for Wakeup Wait

Coefficient	Page	Register	Reset Value	Description
VAD_WW	0x08	0x08	0x01	WW Byte[31:24]
	0x08	0x09	0x38	WW Byte[23:16]
	0x08	0x0A	0x80	WW Byte[15:8]
	0x08	0x0B	0x00	WW Byte[7:0]

Threshold: Threshold (TH) controls the decision boundary of the nodes of the decision tree. A higher value will increase the node thresholds of all the nodes of the decision tree, thus reducing the likelihood of false positives. Similarly, a lower value for the threshold parameter decreases the node thresholds, which reduces the likelihood of false negatives. Thus the threshold parameter can be adjusted to settle at the appropriate balance between false negatives and false positives.

Equation 4 shows the computation of the VAD_TH parameter.

$$Threshold_{new} = Threshold_{default} \times 10^{\frac{thr}{20}} \tag{4}$$

where

- thr is the threshold value in dB (–20 dB to 0 dB)
- $Threshold_{default}$ is the default value in the Threshold register in decimal (16777216)

The default value (16777216) corresponds to 0 dB. Table 2-12 shows the registers that control the VAD_TH parameter.

Table 2-12. Programmable Coefficient Registers for Threshold

Coefficient	Page	Register	Reset Value	Description
VAD_TH	0x09	0x50	0x01	TH Byte[31:24]
	0x09	0x51	0x00	TH Byte[23:16]
	0x09	0x52	0x00	TH Byte[15:8]
	0x09	0x53	0x00	TH Byte[7:0

3 VAD Results

This section discusses the VAD results. The algorithm performance is given by a ROC curve which describes the detection performance across different operating thresholds (–12 dB to –3 dB). ROC plots are included for the noise scenarios from the Aurora Noise database (Figure 3-1 Car , Figure 3-2 restaurant and Figure 3-3 Subway) and speech signals from the NOIZEUS Speech database. Test vectors are generated by mixing noise and speech signals at the desired SNR (SNR is the separation between the power levels of speech and noise signals) of 12, 18, and 24 dB (for example, 12-dB SNR means noise power level is 12 dB down from the speech power level). The operating point is at the extreme top left for the 12-dB threshold, and moves towards the right as the threshold is increased, indicating better performance at Figure 3-4 and the –7-dB threshold for both speech hit rate and non-speech hit rate.
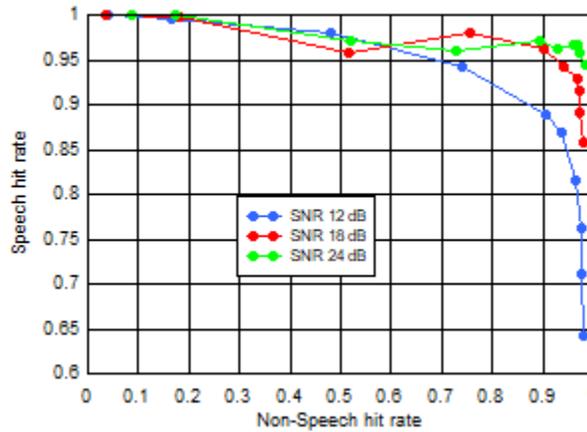
Figure 3-1. Non-Speech Hit Rate vs Speech Hit Rate for Car Noise
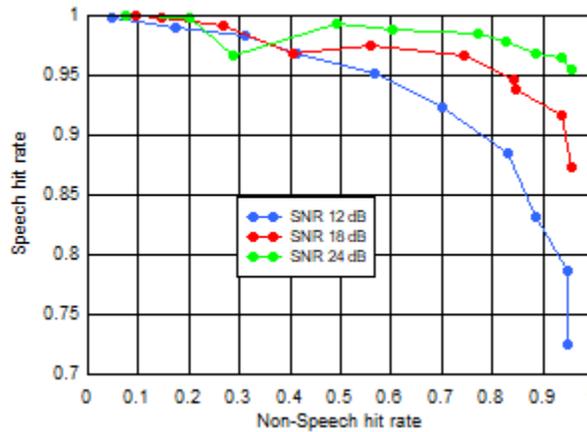
Figure 3-2. Non-Speech Hit Rate vs Speech Hit Rate for Restaurant Noise
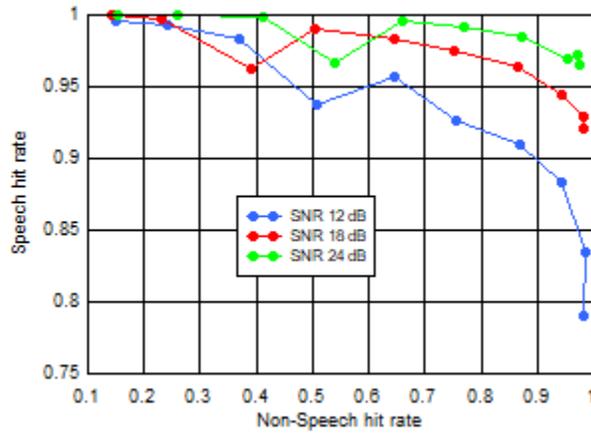
Figure 3-3. Non-Speech Hit Rate vs Speech Hit Rate for Subway Noise

After analyzing the collected data, the –7-dB threshold was chosen to give the best speech hit rate and non-speech hit rate across different noise types. ROC curve at –7-dB threshold for different noise types is as shown.
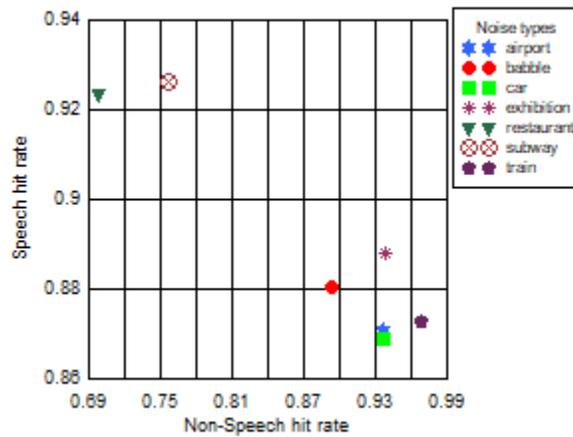

Figure 3-4. Non-Speech Hit Rate vs Speech Hit Rate at –7 dB Threshold for 12 dB SNR

4 Examples

This section presents two examples for configuring the VAD.

Example 1: The following example code shows configuration for using VAD along with analog microphone on CH1.

```
# Key: w 9C XX YY ==> write to I2C address 0x9C, to register 0xXX, data 0xYY
#                # ==> comment delimiter
# See the corresponding EVM user guide for jumper settings and audio connections.
##################################################################
#
#
# Power up IOVDD and AVDD power supplies
# Wait for 1ms.
#
w 9C 00 00 # Goto Page 0
w 9C 02 81 # Wake-up device by I2C write into P0_R2  # Exit Sleep mode
d 10       # Wait for 16 ms
w 9c 00 00 # go to page 9
w 9c 50 00 72 59 dc # -7dB threshold
w 9C 00 00 # Goto Page 0
w 9c 3c a0 # channel 1 set to LINE , SE , AC , 2.5k , DRE disabled
w 9c 73 80 # ip channel 1 enabled
w 9c 74 80 # ASI out ch 1 enabled
w 9c 21 00 # GPIO1 as IRQ (latched only) driving high and low
w 9c 00 01 # go to page 1
w 9c 1e 00 # Channel 1 as VAD input, User mode , internal clk
w 9c 1f 00 # VAD disabled in recording mode
w 9c 00 00 # go to page 0

w 9c 75 21 # Enable VAD
```

Example 2: The following example code shows the configuration for using VAD along with digital microphone on CH1.

```
# Key: w 9C XX YY ==> write to I2C address 0x9C, to register 0xXX, data 0xYY
#                # ==> comment delimiter
# See the corresponding EVM user guide for jumper settings and audio connections.
##################################################################
#
#
# Power up IOVDD and AVDD power supplies
# Wait for 1ms.
#
w 9C 00 00 # Goto Page 0
w 9C 02 81 # Wake-up device by I2C write into P0_R2  # Exit Sleep mode
d 10       # Wait for 16 ms
w 9c 00 00 # go to page 9
w 9c 50 00 72 59 dc # -7dB threshold
w 9C 00 00 # Goto Page 0
w 9c 3c 40 # Ch1 configured as Digital microphone PDM input
w 9c 41 40 # Ch2 configured as Digital microphone PDM input

w 9c 73 C0 # ip channel 1 and channel 2 enabled
w 9c 74 C0 # ASI out ch 1 and ch 2 enabled
w 9c 21 21 # GPIO1 as IRQ
w 9c 22 41 # GPO1 as PDM CLK
w 9c 2b 45 # gpi1 as Ch1/2 DMIC data
w 9c 20 00 # gpi1 neg edge as ch1

w 9c 00 01 # go to page 1
w 9c 1e 00 # Channel 1 as VAD input, User mode , internal clk
w 9c 1f 00 # VAD disabled in recording mode
w 9c 00 00 # go to page 0


w 9c 75 21 # Enable VAD
```

5 Related Documentation

For related documentation see the following:

- TLV320ADC6120
  - Texas Instruments, *TLV320ADC6120 2-Channel, 768-kHz, Burr-Brown^{TM} Audio ADC data sheet*
  - Texas Instruments, *TLV320ADC6120 stereo-channel, 768-kHz, Burr-Brown™ audio ADC with 106-dB SNR evaluation module*
- TLV320ADC5120
  - Texas Instruments, *TLV320ADC5120 2-Channel, 768-kHz, Burr-Brown^{TM} Audio ADC data sheet*
  - Texas Instruments, *TLV320ADC5120 stereo-channel, 768-kHz, Burr-Brown™ audio ADC with 106-dB SNR evaluation module*
- TLV320ADC3120
  - Texas Instruments, *TLV320ADC3120 2-Channel, 768-kHz, Burr-Brown^{TM} Audio ADC data sheet*
  - Texas Instruments, *TLV320ADC3120 stereo-channel, 768-kHz, Burr-Brown™ audio ADC with 106-dB SNR evaluation module*
- Texas Instruments, *PCM6120-Q1 2-Channel, 768-kHz, Burr-Brown^{TM} Audio ADC* data sheet.
- Texas Instruments, *PCM5120-Q1 2-Channel, 768-kHz, Burr-Brown^{TM} Audio ADC* data sheet.
- Texas Instruments, *ADCx120EVM-PDK User's Guide*
- Texas Instruments, *PurePath™ Console*

6 Revision History

Changes from Revision * (December 2021) to Revision A (April 2022) **Page**