VocalFusion® XVF3620 Voice Processor

AI-enhanced voice capture for communication and speech recognition with ultra-fast convergence

XMOS

A reliable, fast time-to-market, and cost optimised solution for voice capture even in noisy and changing environments for communication, speech recognition, LLM interfacing & robotics.

The VocalFusion® XVF3620 is a high-performance voice processor designed for reliable voice control and communication in noisy and reverberant environments, including those with non-stationary noise such as machinery and outdoor environments.

It integrates a complete on-chip digital voice processing pipeline with AI-based noise reduction and fast-adapting acoustic echo cancellation. Using two MEMS PDM microphones, it captures and processes audio to deliver high-quality output for speech recognition or human communication, with AI tuning adaptable to specific use cases, such as storms, clicks and dog barking.

Key features include AI denoising, full-duplex acoustic echo cancellation with residual filtering, two-microphone beamforming, and automatic gain control — improving speech clarity, reducing background noise, and stabilising audio levels without external processing.

The device supports flexible integration with I2S or USB audio routing and USB or I2C control. It operates in either USB Accessory mode (USB Audio Class 2.0 for audio and control) or Integrated mode (I2S audio with I2C control).

The XVF3620 is ideal for consumer electronics and industrial robotics requiring robust, real-world voice performance.

FEATURED HIGHLIGHTS

AI-BASED NOISE REDUCTION

Background noise is reduced using a trained AI model that preserves speech while suppressing unwanted sounds. This enhances audio clarity, even in noisy and dynamically changing environments. The processing can be tuned for speech recognition or voice communication, making it well suited to non-stationary noise scenarios such as a storm affecting a door intercom.

FAST-ADAPTING ACOUSTIC ECHO CANCELLATION

Full duplex echo mono cancellation is applied to remove loudspeaker echo from the microphone signal. Acoustic models are adapted continuously to follow room and movement changes. Reliable barge-in and stable performance are maintained.

TWO-MICROPHONE VOICE CAPTURE WITH BEAMFORMING

- 32-bit floating-point scalar pipeline offering up to 1200 MFLOPS of performance at 600MHz
- 256-bit VPU adds block floating point capabilities up to 38.4 GMACC/S of performance at 800MHz
- Integrated complex arithmetic and FFT/iFFT support at up to 1 million 256-point FFT/s

FLEXIBLE SYSTEM INTEGRATION

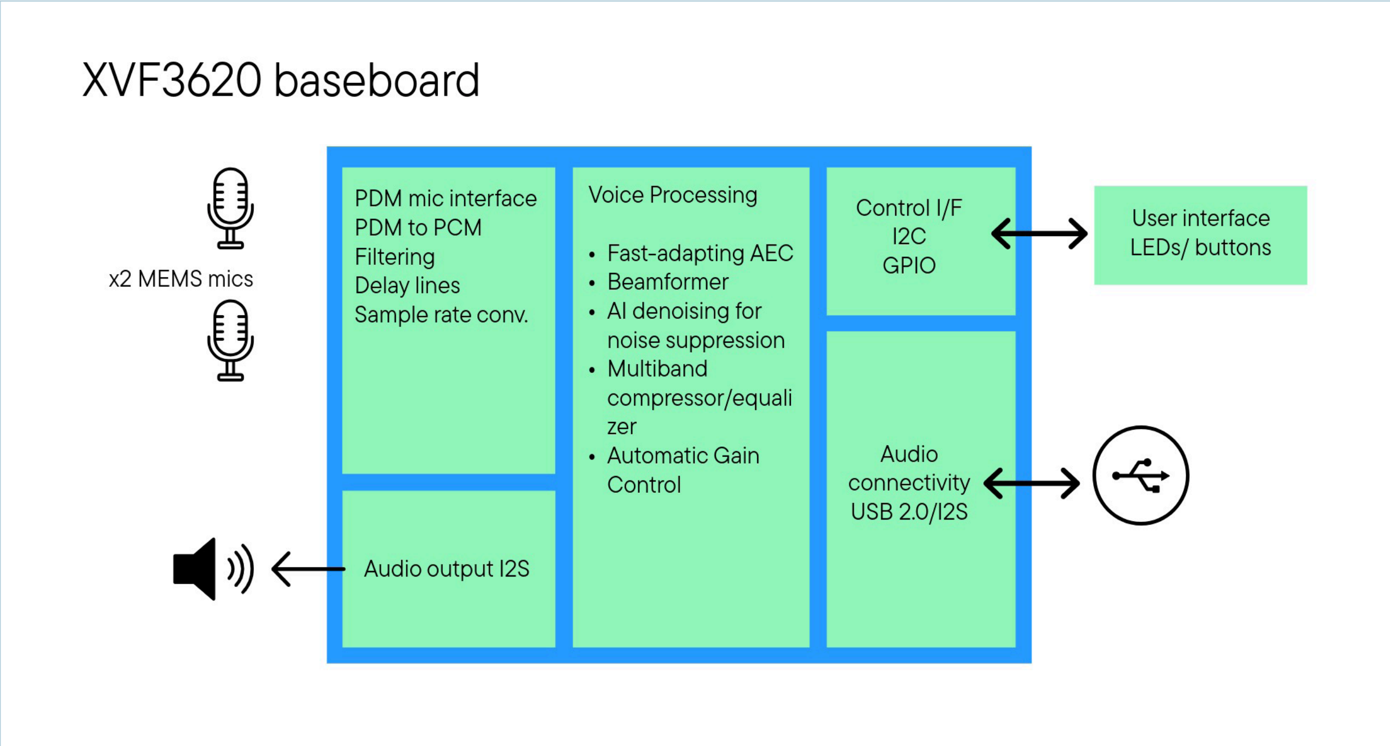
A complete voice processing pipeline is integrated on the device. Audio and reference signals are supported over I2S or USB, with control via USB or I2C. USB Accessory and Integrated operating modes are supported for different product designs.

APPLICATIONS

- INTERCOM SYSTEMS
- OUTDOOR ENVIRONMENTS
- SMART HOME APPLIANCES
- SMART ROBOTS
- INDUSTRIAL ASSISTANTS

BLOCK DIAGRAM

XMOS



VOICE PROCESSOR - XVF3620

PACKAGE
60-Pin QFN, 0.4mm pitch

VOICE PROCESSING
1-2x Microphone PDM to Pulse Code Modulation (PCM) conversion
Fast-adapting acoustic echo cancellation
Beamformer (GSC)
AI denoising
Multiband compressor/equalizer (MC/EQ)
Automatic Gain Control (AGC)

MICROPHONE INTERFACE
2x digital PDM interface
100mm linear mic array, 33mm inter-mic spacing
90mm square mic array, 43mm inter-mic spacing
Mic spacing up to 100mm

HOST INTERFACE OPTIONS
High speed USB2.0 device supporting USB Audio
USB Class 1 or Class 2
Class 2.0; 16kHz or 48kHz sample rate
I2S audio interface; 16kHz or 48kHz sample rate

AUDIO OUTPUT OPTIONS
I2S output to DAC; 16kHz or 48kHz PCM

CONTROL INTERFACE
USB Control Interface
I2C Control Interface

DEV KIT - XK-VOICE-L71