A Guide to Developing Voice- enabled Products using i.MX RT600 MCUs

March 2021

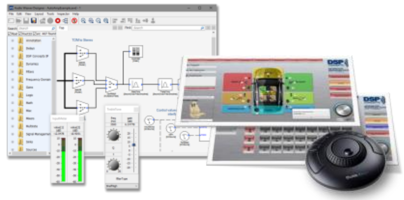
Agenda

- DSP Concepts Intro & Overview
- Voice-enabled Product Design Guidelines
  - Voice Stack Architecture and Overview
  - Microphone Considerations
  - Loudspeaker Considerations
  - System Considerations
  - Recommended Designs
- TalkTo Demo

DSP Concepts Intro and What We Do?

- Founded in 2003 as consultancy, transitioned to a product company in 2014
- We empower industry with a standardized audio processing platform & IP

GoPro TESLA logitech BMW BRAUN PORSCHE



Audio Weaver Platform

- Audio Weaver Designer
  - Graphical Audio Development Tools
- Audio Weaver Core
  - Embedded Processing Engine
- Real Time Audio System Check
  - System Validation

TalkTo

Voice Comms

PlayPack

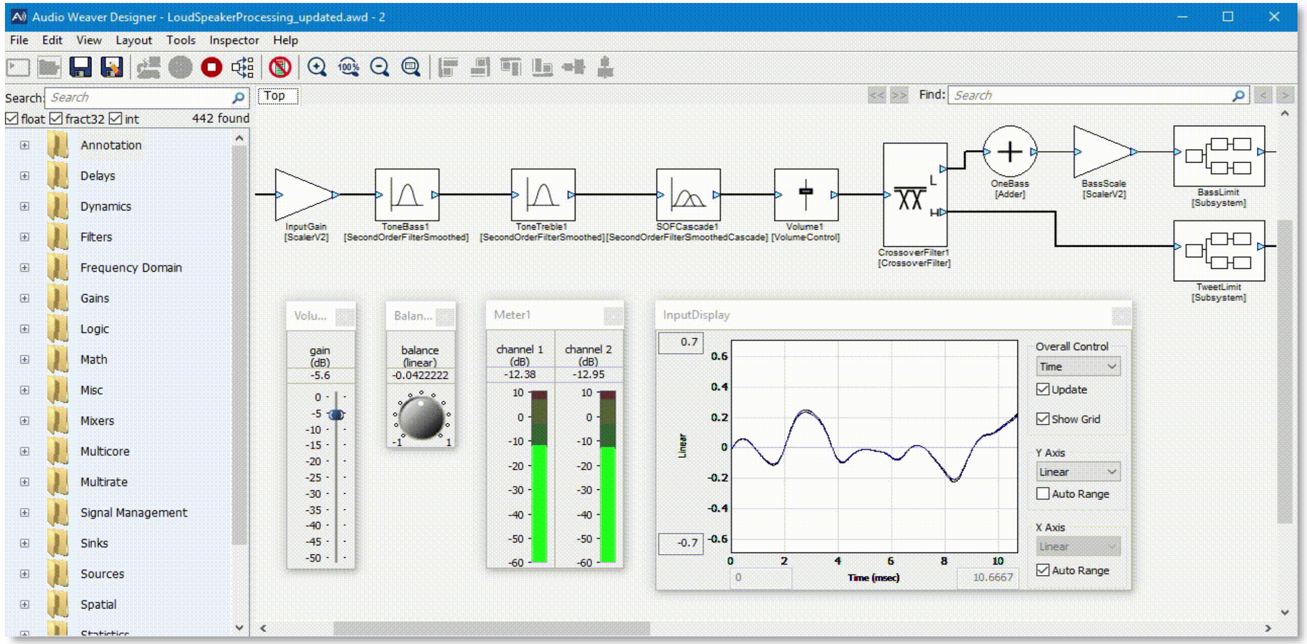
Audio Weaver Solutions

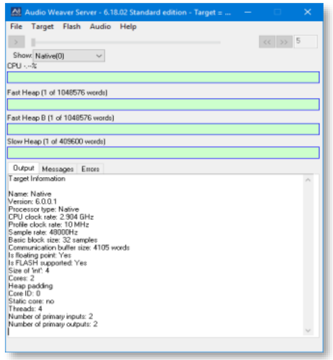
- TalkTo Audio Front End
  - Supports many form factors
- Voice Communications
  - Full Duplex Voice Calls
- PlayPack Audio Playback Suite
  - **PlayVoice** – Improves voice intelligibility
  - **PlayBass** – Improves low frequency response
  - **PlayLevel** – Normalizes volume
  - **PlayWide** – Increases perceived sound stage width

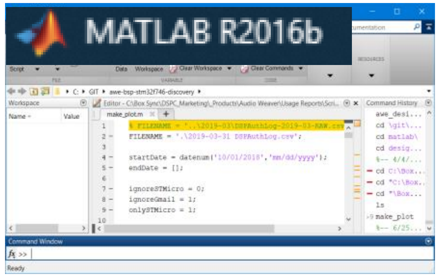
Engineering Services

- System Architecture
- System profiling & check (RTASC)
- Mic-array Design
- Loudspeaker tuning
- Audio Weaver Custom Modules
- Custom Audio Solutions

Audio Weaver: Design, Tune, and Deploy







Audio Weaver Designer

Configuration tools for Audio Weaver Core

- Integrate & Create advanced audio features
- Real-time Interface for design and debug
- Open APIs for external tools and scripting

Audio Weaver Core

All-in-one Audio Processing Engine

- Runtime-reconfigurable Audio-Pipeline
- 400+ Audio Building-blocks
- Runs on MIMXRT685-EVK Board
    - Optimized for Cadence® Tensilica® HiFi 4 DSP

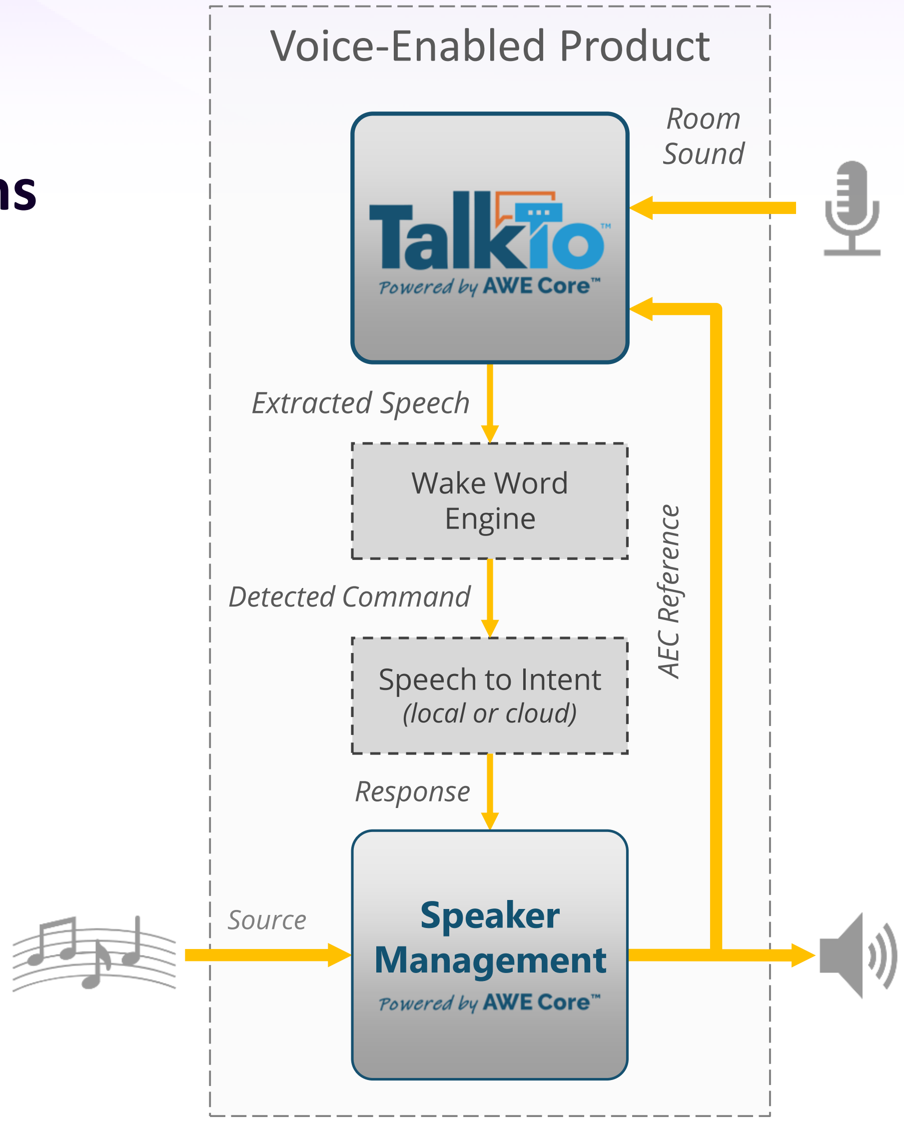
TalkTo on i.MX RT600 MCUs

TalkTo Features

- Exceptional far field performance
- Mono, Stereo, or multichannel AEC
- Adaptive Interference Canceller™
- 2, 4, or 6 mic geometries
- Custom mic-arrays supported
- Optional loudspeaker processing
- Optimized for i.MX RT685 MCU

Targeted Applications

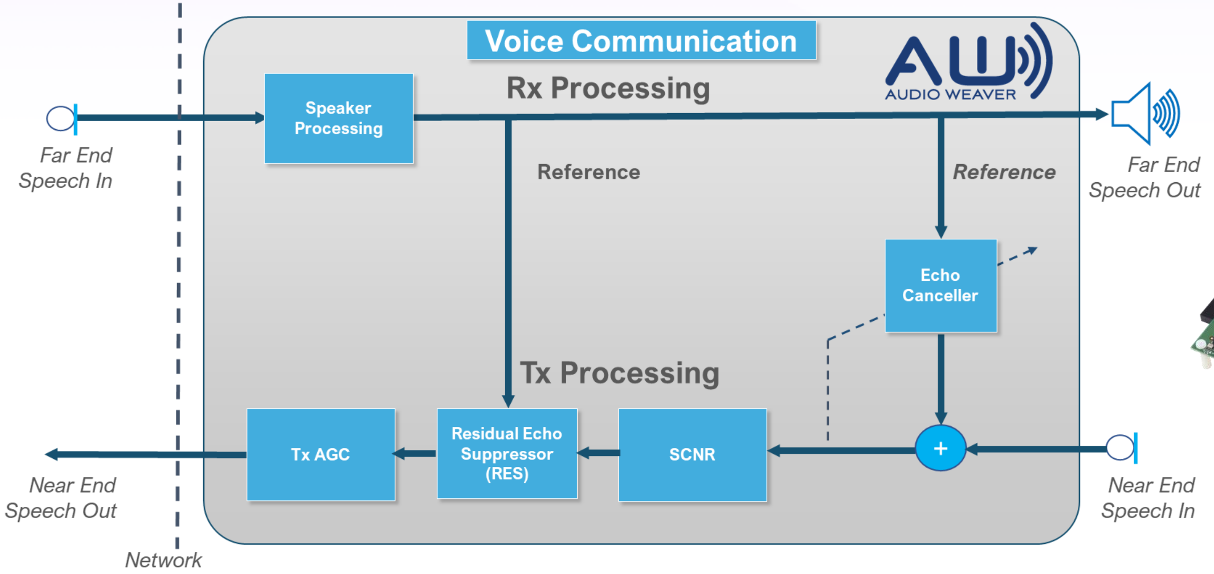
- IoT & Smart Home
- Appliances
- Smart Speakers
- TVs & Set-top boxes
- Multichannel Soundbars



Voice Communication on i.MX RT600 MCUs

Voice Comm Features

- Wideband, high-quality voice
- Full duplex operation
- Automatic Distance Compensation
- Acoustic Echo Cancellation
- Active noise-reduction
- De-reverberation
- Comfort Noise
- Third-party VOIP clients
- Can combine with TalkTo
- Optimized for i.MX RT685 MCU



Targeted Applications

- IoT & Smart Home
- TVs & Set-top boxes
- Whole-house Intercom
- Medical
- Speakerphone
- Office / Commercial
- Industrial

PlayPack on i.MX RT600 MCUs

PlayPack

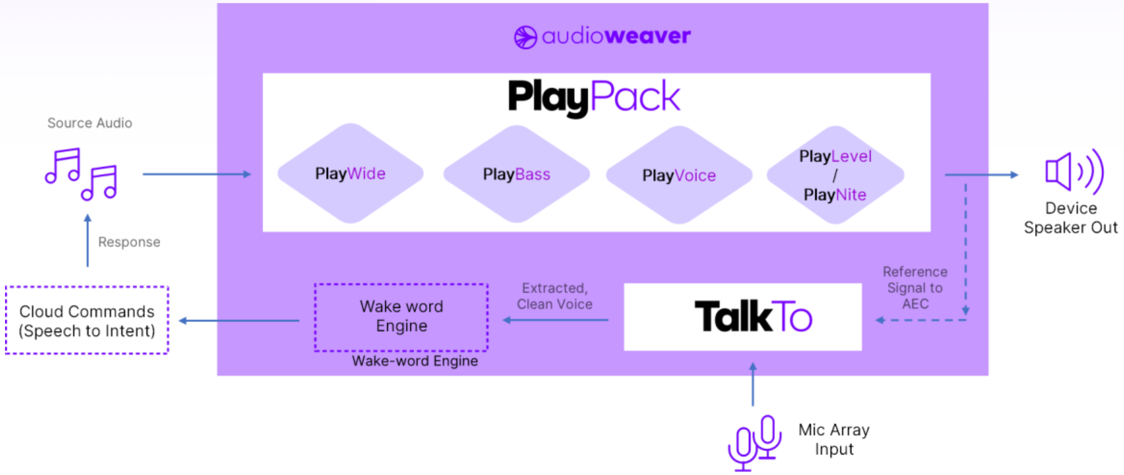
PlayBass
Bass enhancement. Perceptual algorithm for improving small speaker performance.

PlayLevel
Volume management. Normalizes content for uniform listening experience. Eliminate annoying loud transitions.

PlayWide
Spatialization. Increases perceived width of sound stage.

PlayVoice
Dialogue enhancement. Pulls vocals out of the mix.

PlayNite
Dynamic compression for minimizing special effects



PlayPack Features

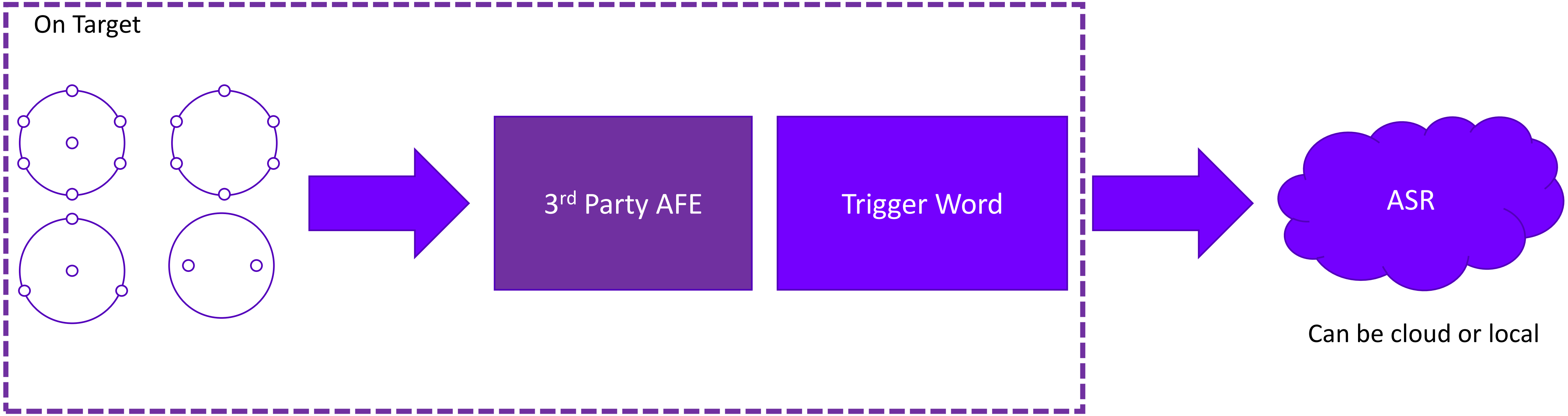
- Enhance low-cost loudspeaker performance
- Tunable
- Combine with TalkTo and TalkTogether
- Source files for customization

Targeted Applications

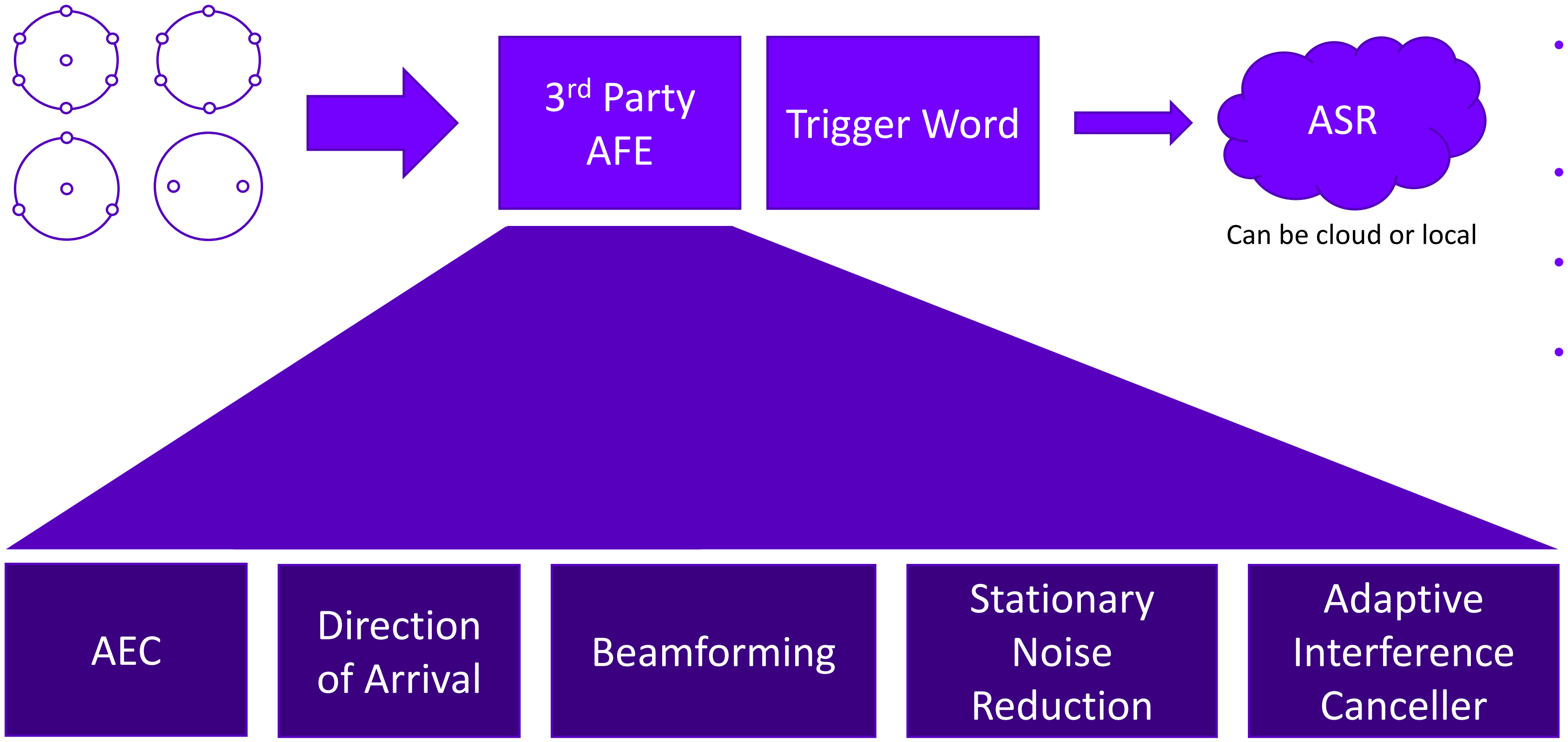
- Bluetooth Speakers
- Laptops
- Radios
- Smart Speakers
- Smart TV
- Soundbar

Voice-enabled Product Design Guidelines

Typical Voice Stack



Looking closer at the AFE



- **Acoustic Echo Cancellation (AEC) –** allow voice commands to be heard while music is being played
- **Direction of Arrival –** determines location of sound source that helps steer the beamformer
- **Beamforming –** combine multiple microphone signals to improve SNR
- **Noise Reduction –** models background noise and then removes
- **Adaptive Interference Canceller –** combines beamforming statistical signal processing, and machine learning to reduce directional noise

Comparing Amazon versus Google

- 2 microphones only
- 65 to 71mm spacing
- Mono or stereo
- High-end application processor required

- No variation in products
- No variation in performance
- Performance lags behind AVS

- Any number of microphones
- Any spacing
- Any number of playback channels
- Application processor MCU solutions

- Wide variety of designs
- 2 to 7 microphones
- Different form factors
- Better performance
- Low-cost designs possible

Microphone Considerations – Microphone Types

Analog

- Outputs an analog differential signal
- Requires an external A/D converter
- More difficult to interface
- Prone to interference noise problems

Digital MEMS

- Output a digital bitstream in PDM format
- Processor requires a PDM to PCM converter
- Easier to interface
- Easier to achieve low noise since inherently digital interface
- Preferred by product makers

Microphone Considerations – Data Sheet Specs

Symbol	Parameter	Test Condition	Min.	Typ.	Max.	Unit
Vdd	Supply Voltage		1.64	1.8	3.6	V
Idd	Current Consumption in normal mode	Mean value		0.6		mA
IddPdn	Current consumption in power-down mode			20		uA
Scc	Short-circuit current		1		10	mA
AOP	Acoustic overload point			120		dBSPL
So	Sensitivity		-29	-26	-23	dBFS
SNR	Signal-to-noise ratio	A-weighted at 1kHz, 1 Pa		65		dB

Microphone Considerations – Levels

Acoustic Levels

- Measured in dB SPL = “decibels of sound pressure level”
- 94 dB SPL is the reference level used throughout the industry. Equals 1 Pascal.

SPL meters measure this

Digital Levels

- Measured in dB FS = “decibels relative to full scale”
- Maximum level is 0 dB = 1.0

Digital levels measured in software

Connects digital and acoustic levels

Microphone Considerations – Sensitivity

Symbol	Parameter	Test Condition	Min.	Typ.	Max.	Unit
AOP	Acoustic overload point			120		dBSPL
So	**Sensitivity**		**-29**	**-26**	**-23**	**dBFS**
SNR	Signal-to-noise ratio	A-weighted at 1kHz, 1 Pa		65		dB

94 dB SPL → -26 dB FS

Sensitivity has min, max, and typical values.
Good microphones have +/-1 dB tolerance (you have to pay extra).
Typical microphones have +/- 3 dB tolerance.

Microphone Considerations – AOP

Symbol	Parameter	Test Condition	Min.	Typ.	Max.	Unit
AOP	**Acoustic overload point**			**120**		**dBSPL**
So	Sensitivity		-29	-26	-23	dBFS
SNR	Signal-to-noise ratio	A-weighted at 1kHz, 1 Pa		65		dB

- Specifies at what SPL level the microphone clips
- Usually corresponds to the 0 dB FS point (when does the digital signal clip?)

Microphone Considerations – SNR

Symbol	Parameter	Test Condition	Min.	Typ.	Max.	Unit
AOP	Acoustic overload point			120		dBSPL
So	Sensitivity		-29	-26	-23	dBFS
SNR	**Signal-to-noise ratio**	**A-weighted at 1kHz, 1 Pa**		**65**		**dB**

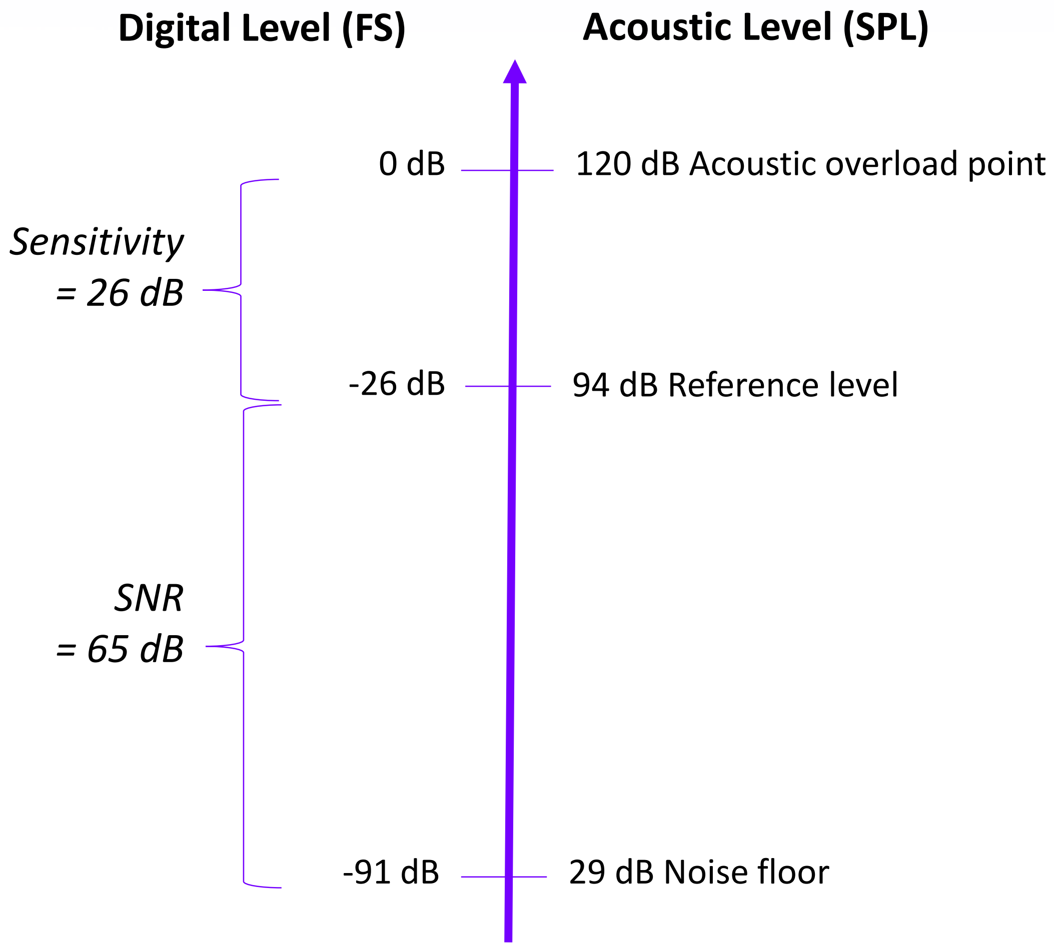
- SNR is relative to the 94 dB reference level of the microphones
- This allows you to determine the digital noise floor
- 65 dB SNR sufficient for most applications → higher SNR helpful with smaller spacing

94 dB SPL → -26 dB FS   Reference level

↓ 65 dB SNR

29 dB SPL → -91 dB FS   Noise floor

Microphone Considerations – Combined Specs



Typical speech level is 65 dB SPL

This would correspond to a digital level of -55 dBFS with this microphone

Microphone Considerations – Acoustical Porting

(No Common Cavity)

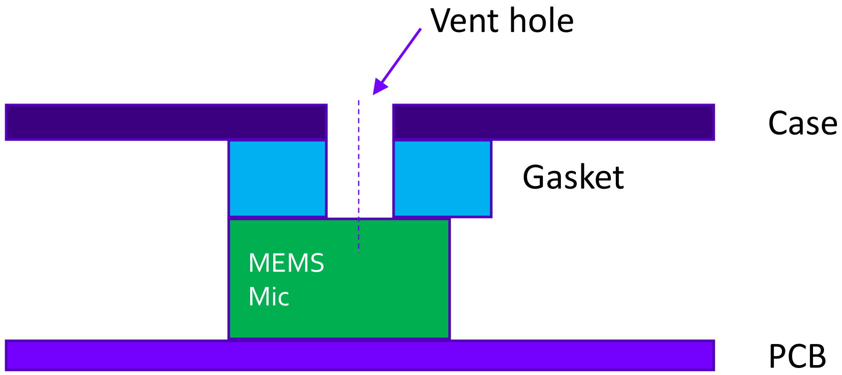
- This design with a common cavity shared by all microphones will not work

- You need individual gaskets to make a direct connection between each mic and its vent hole

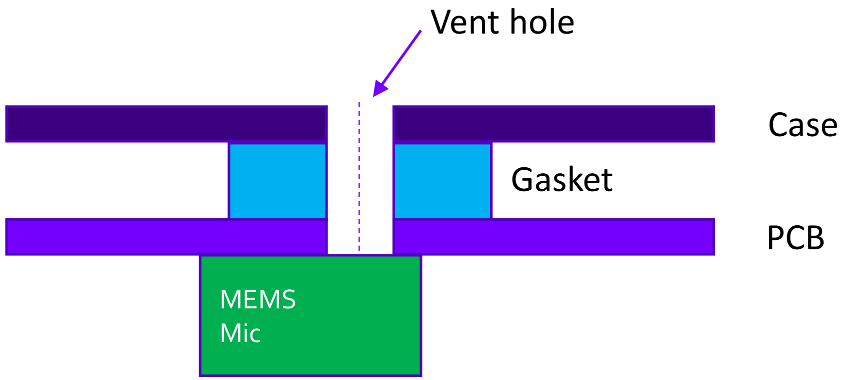
Microphone Considerations – Acoustical Porting

(Minimize Port Length)

Top Ported (recommended)



Bottom Ported (not as good)



$$f_b = \frac{c \times D}{4\sqrt{\pi \times V \times (L + \sqrt{\pi} \times D/2)}}$$

where:

f_b is the resonance frequency, Hz.

c is the speed of sound, approximately 340 m/sec.

D is the vent diameter, mm.

V is the cavity volume, mm^3.

L is the vent length, mm.

Helmholtz resonance

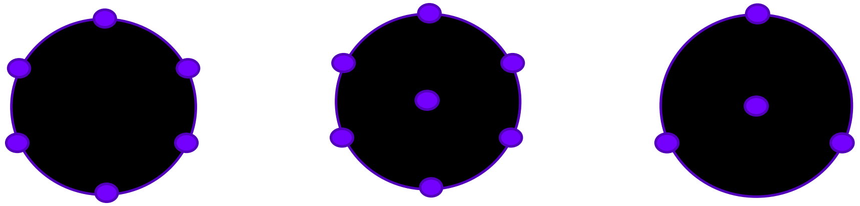
- Make sure the resonance is outside of voice band

- The top ported design has a shorter vent length L
- This shifts the resonance higher in frequency out of the voice band.

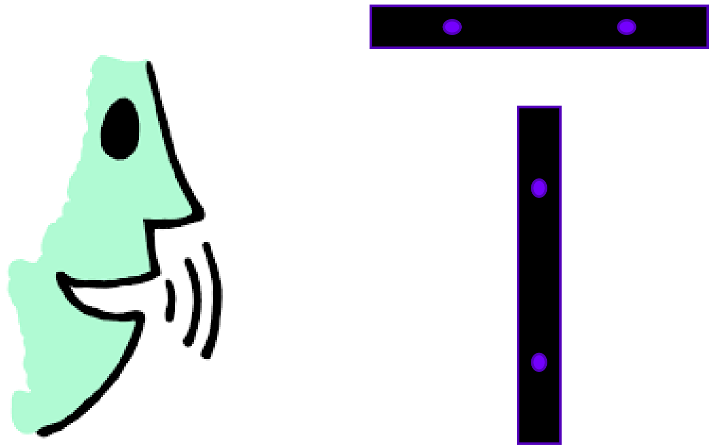
Microphone Considerations – Arrays

- Circular microphone arrays preferred
  - Supports 180 or 360 degree operation
  - Provides 5 – 7 dB SNR improvement
- Linear arrays work well when in an end-fire configuration
  - Requires person to be in a specified location
  - Provides 4 to 5 dB SNR improvement
- Broadside arrays work poorly and should be avoided
  - Very little SNR improvement to low frequencies where the bulk of speech energy resides
- Use broadside arrays only as a last resort when the industrial design dictates no other options
  - Television
  - Wall Panel

Circular Arrays



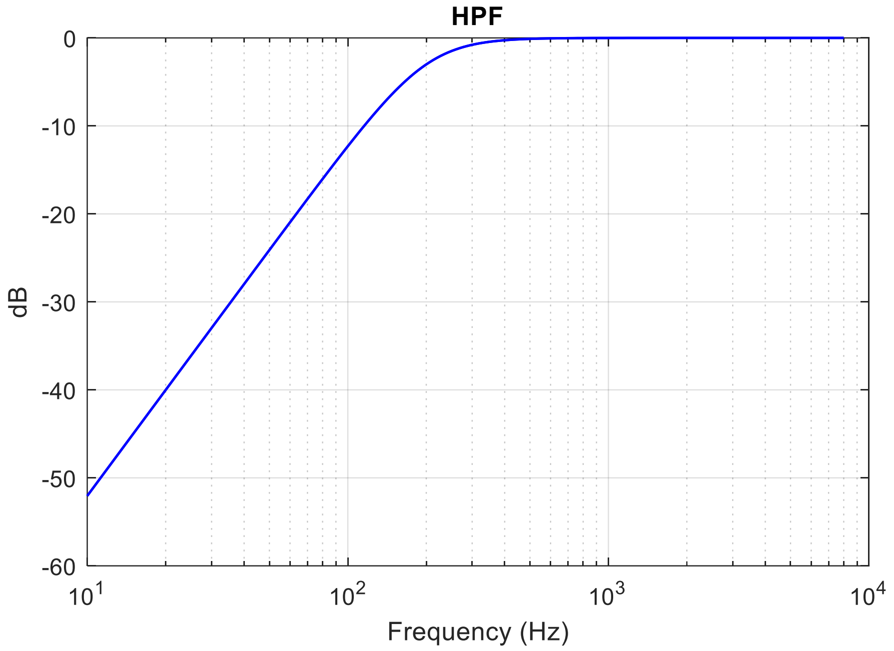
End-fire

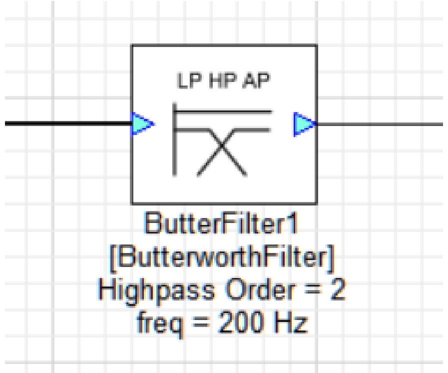


Broadside

Microphone Considerations – High Pass Filter

- Always put a high pass filter at the start of your system
- Eliminates low frequency noise
- Removes DC offset from microphones
- Typical voice range 200 Hz – 8 kHz
- Recommendation: 2nd order Butterworth high pass filter with a cutoff of 200 Hz



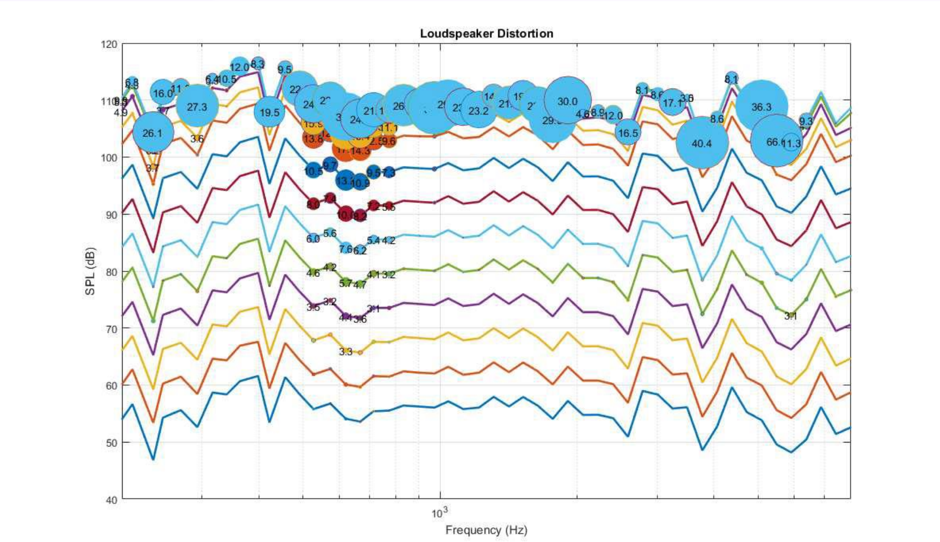


Microphone Considerations - Summary

- Microphones should be:
  - placed on top of product if possible
  - on a flat horizontal surface
  - visible to the user
- SNR ~ 65dB
- Gains matched to within +/- 1dB
- Microphone AOP must be high enough so that the system doesn't clip when loudspeakers are played at full volume. Recommendations:
  - 120 dB for smart speakers
  - 130 dB for sound bars
- Recommend MEMS microphones (not ECM) since they are better matched.
- Circular microphone arrays preferred over linear microphone arrays
- 70mm microphone spacing recommended but can go down to 40mm with some performance degradation

Loudspeaker Considerations

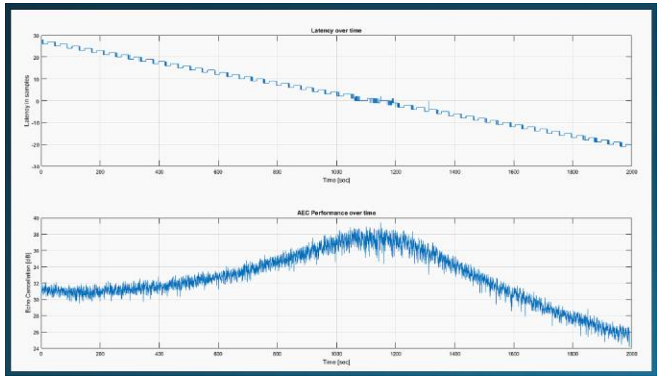
- **Loudspeaker-to-microphone coupling**
  - Coupling through air and enclosure
  - Maximize distance and seal appropriately
- **Distortion**
  - Limiting factor for AEC
  - Cabinet resonances and internal nonlinearities
  - Measure across all frequencies
  - Low frequencies usually the issue
  - Compensate reference channel

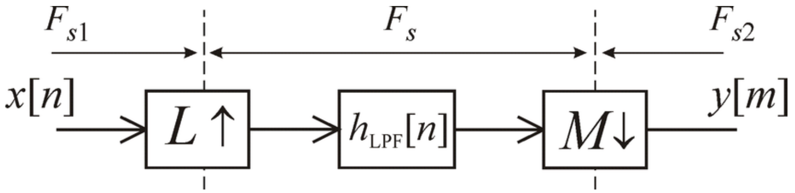


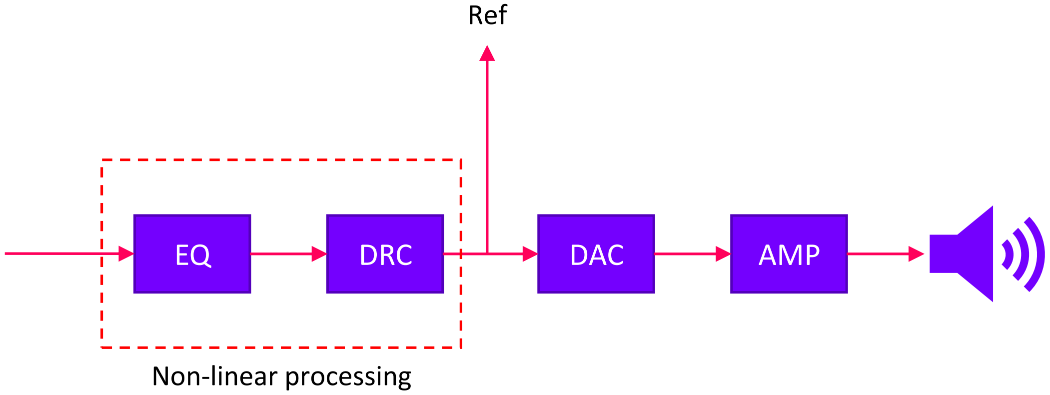
Loudspeaker THD	Max AEC Cancellation
1%	40 dB
2%	36 dB
5%	26 dB
10%	20 dB
20%	14 dB

System Considerations

- **Reference Signal to Microphone Latency**
  - Ensure latency is stable over time
  - Jitter +/- 1 sample is OK due to measurement noise
  - Large discontinuities or ramping delays are a problem
- **Sampling Rate**
  - Use same clocks for audio streams whenever possible
  - Use Asynchronous Sample Rate Converter (ASRC) for independently clocked audio streams
- **Speaker Processing**
  - Reference signal must be taken **after** nonlinear processing
  - Separate nonlinear processing requires 2 AEC reference channels
  - Watch out for smart amplifiers



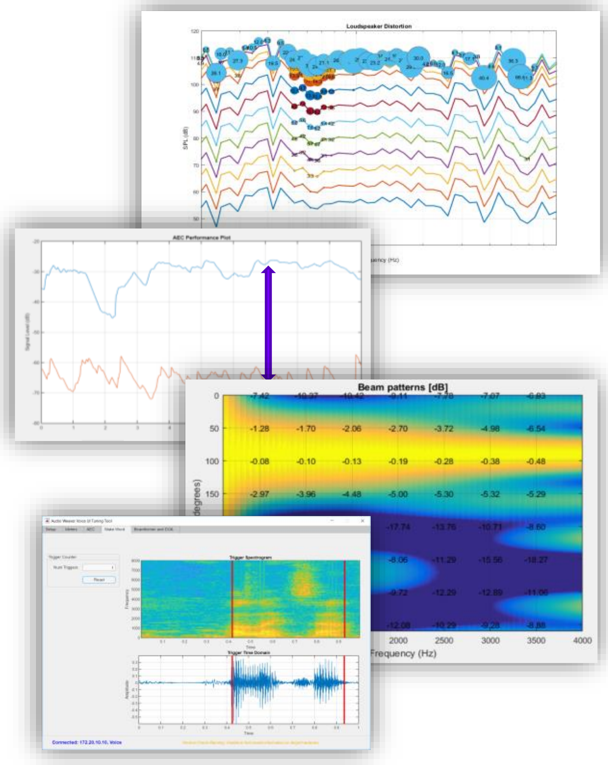




RTASC can help with hardware considerations

RTASC = Real-Time Audio System Check

Enable in form factor real-time debugging and validation of audio/voice hardware, components, and systems.



Typical Uses

- Debug acoustics in development
- Validate end product/component audio performance during manufacturing
- Validate semiconductor performance before tape out

Microphones

- Noise floor
- Acoustic overload point
- Spurious harmonics
- High frequency rejection
- Isolation / porting
- Conducted sound
- Gain mismatch
- Time synchronization

Distortion / THD

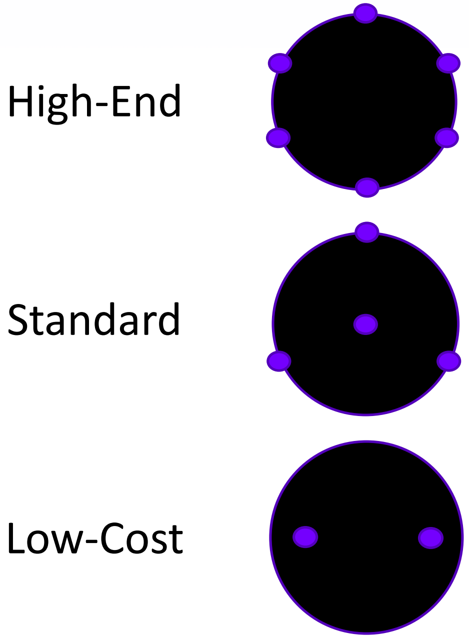
- Loudspeaker nonlinearities
- Coupling between loudspeakers
- and microphones
- Enclosure resonances

Audio Clocking

- Mic and playback clock stability
- Fixed latency through system
- Audio buffer over and underruns
- Dropped samples

Recommended Product Designs

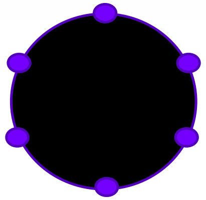
Smart Speaker Designs



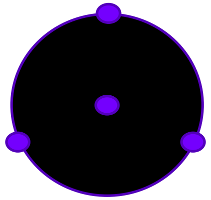
- 360-degree operation
- Microphones on top of product
- 40 to 75 mm diameter
- Physically separate microphones and loudspeakers for best performance
- Mono or stereo playback

Sound Bar Designs

High-End

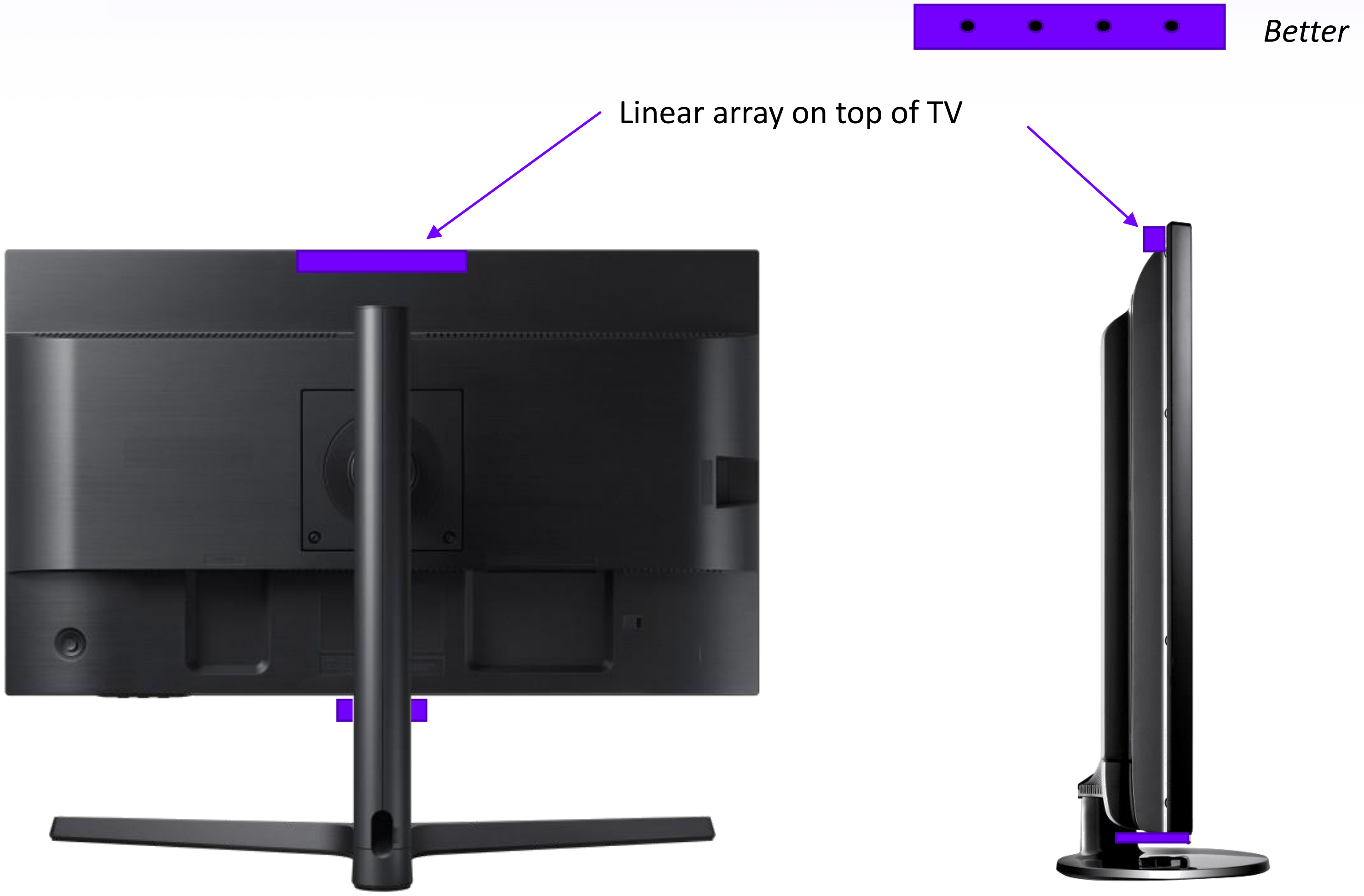


Standard



- 180-degree operation
- Microphones on top of product near center of device
- 60 to 75 mm design
- Physically separate microphones and loudspeakers for best performance
- Stereo or multichannel playback (up to 7 reference channels)
- Compatible with Dolby Atmos

TV Designs

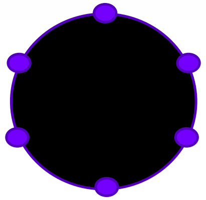


Placement options

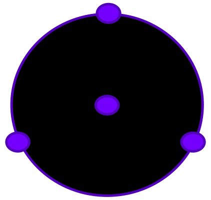
- Better performance
  - 4 mic linear array on top
  - Microphones facing up or front
- Good performance
  - 2 mic linear array on top
  - Microphones facing up or front

Set Top Box Designs

High-End



Standard



- Top of Device
- 180-degree operation
- Microphones on top of product
- Tethered “puck”
- 360-degree operation
- Microphones on top of product

- Support for optional internal speaker for voice playback
- Audio playback through HDMI

Appliance / Tablet Designs

Good

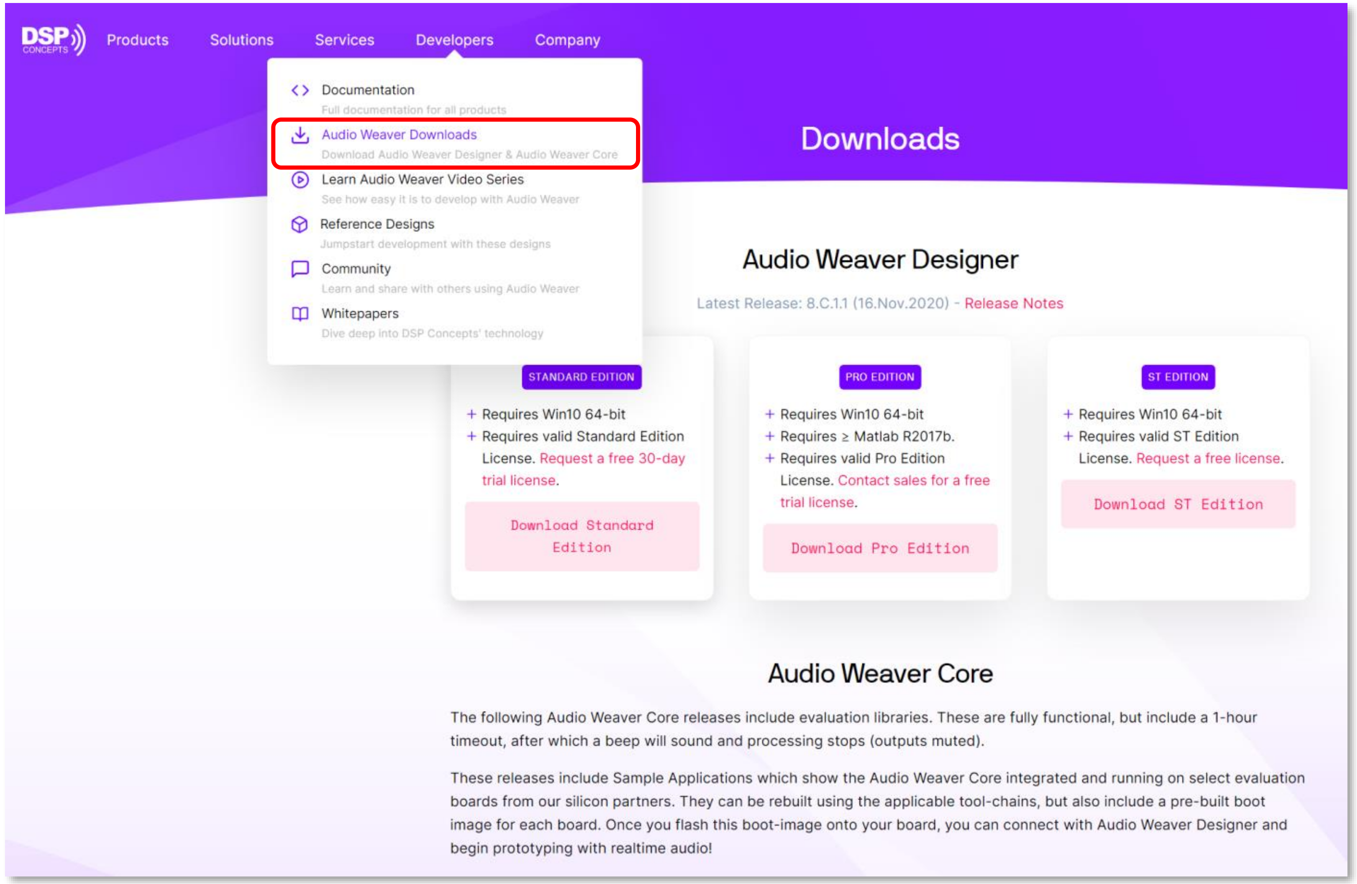
Better

- 180-degree operation
- 2 or 4 microphone linear array
- 25 to 75 mm design
- Physically separate microphones and loudspeakers for best performance
- Mono or stereo playback

TalkTo Demo

Next Steps: Evaluate Audio Weaver & Solutions

1. **Free 30-day License for Audio Weaver Designer:**



https://dspconcepts.com/audio-weaver-trial

2. **MIMXRT685-EVK Development Board ($129 USD) available from NXP and its distributors**

i.MX RT600 Device family:
https://www.nxp.com/imxrt600

3. **Contact DSP Concepts for Audio Weaver Core and/or TalkTo Development Licenses**

Follow-up Questions
Mike Vartanian
mike.vartanian@dspconcepts.com

audioweaver

DSP CONCEPTS