Microsoft

Windows 11 (SV2)

AI Experiences on NPU – OEM Enablement Guide

Version 1.1

Abstract

This document provides OEMs with the technical guidance to enable key AI experiences leveraging hardware acceleration on NPUs in the Windows 11 2022 Update (codename Sun Valley 2 or SV2) OS release.

Information is subject to change.

Feedback and questions should be directed to your key Microsoft account and co-engineering representatives, which will help us to provide updates in future versions of this document.

Confidentiality Reminder

Microsoft values our deep co-engineering partnership with key silicon IHVs delivering NPUs. Documents shared within the 'AI Experiences' engagement on Microsoft Collaborate have been carefully structured to protect and respect the confidentiality of our IHV partners.

We understand that close collaboration between Microsoft, OEMs, ODMs, and IHVs is needed to successfully enable these deeply integrated platform capabilities and experiences in Windows. Any leak of this confidential information could seriously impact on Microsoft's product plans and the plans of our co-engineering IHVs.

When referring to information contained within these documents, please maintain confidentiality by following these principles:

- Documents published on Microsoft Collaborate are the main source of up-to-date information being shared across OEM, ODM, IHV partners. When discussing details contained within these documents with relevant parties, please refer to the link to the package URL on Microsoft Collaborate to safeguard confidentiality and to ensure that the latest information is available across companies.
- Partners with 'need to know' access must be properly added to the respective 'AI Experiences' engagements on Microsoft Collaborate. Please work with your Microsoft representative for access permissions.
- Do not share content pertaining to Windows AI investments outside of Microsoft Collaborate. This includes copying documents to internal repositories (e.g. Teams sites, SharePoint, Cloud storage) that are accessible to multiple partners, attaching documents to e-mail, or copying and pasting content into e-mail.
- When collaborating with a silicon IHV, do not share IHV-proprietary information from one IHV to another IHV. Documentation has been purposely structured, compartmentalized, and permissioned to prevent cross-IHV leaks of information. In multi-party meetings with an IHV, do not discuss plans or features of other IHVs which would violate their confidentiality.

Thank you for your support in maintaining the confidentiality of our plans and our IHVs.

Revision History

Version	Date	Key Changes
0.1	1/31/2022	Initial version
0.2	3/7/2022	Updates to Camera AI Effects co-existence, Added Validation/HLK Updates to Microsoft Effect Pack Opt-in Added Voice Focus enablement content Added Live Captions enablement content
0.3	6/22/2022	Added additional information on Camera opt-in. Add clarification of Camera AI effects opt-in on legacy camera device interfaces. Added more details on how on Camera default settings work. Added update on profile-aware apps now having the capability to access photo pin with full resolution. Support available in Windows 11 SV2 (22H2) 7D servicing release.
1.0	2/8/2023	Added confidentiality reminder Add details on Windows Studio taskbar Quick Settings Added revisions to testing process and sequence for Voice Focus. Corrected details about camera profiles, DDI support and MediaTypes under "Key System and Camera points to note for MEP Opt-In" and "Post Opt-in Expectations". Added section on "Leveraging higher resolution processing in MEP" Added DEVPROPKEY for MEP camera to run in "high resolution mode" Clarified expectations for HLK under "Validating Camera Effects" Removed Voice Focus validation requirements. Details covered in separate guidance documentation to partners enabling this feature.
1.1	11/20/2023	Added links to additional DMFT resources under "Design considerations for OEM effects extensibility and co-existence. Updated the guidance for WSE camera opt-in.

Table of Contents

Glossary

Acronym	Term
AEC	Acoustic Echo Cancellation
CER	Customer Experience Review
CI	Continuous Innovation
DDI	Device Driver Interface
DMFT	Device Media Foundation Transform
HLK	Hardware Lab Kit
M365	Microsoft 365
MEP	Microsoft Effect Pack
ML	Machine Learning
NPU	Neural Processing Unit
RTC	Real-Time Communication
SoC	System on Chip
SV2	Sun Valley 2
UVC	USB Video Class
WHCP	Windows Hardware Compatibility Program
WSE	Windows Studio Effects

Introduction

This document provides OEMs with the technical guidance to enable key AI-powered experiences leveraging hardware acceleration on NPUs in the Windows 11 22H2 OS Release (publicly referred to as the Windows 11 2022 Update and codename Sun Valley 2 or SV2). Information within this document is also applicable to subsequent OS updates based on the Windows 11 22H2 platform, including Continuous Innovation releases in CY2023:

- 2023 February CI
- 2023 May CI
- 2023 Sept CI

This document focuses on the hardware dependent NPU experiences that require OEM enablement including:

- **Windows Studio Effects (also known as Microsoft Effect Pack) camera and audio effects:**
  - Camera AI effects (Background Blur, Eye Contact, Automatic Framing)
  - Deep Noise Suppression (Voice Focus)
- **Natural and Intuitive Speech:** Live Captions, Voice access

This information in this document is subject to change. Where appropriate, this document will refer to separate documents containing IHV-specific information to maintain IHV confidentiality. References to Windows Studio Effects, Microsoft Effect Pack, or MEP are used interchangeably within this document.

Vision

AI-Infused Experiences and Devices

Windows 11 aims to deliver smarter, more intuitive, more productive, and delightful experiences to customers through a combination of Windows devices, Microsoft applications, and services that are infused with AI. This includes a broad strategy to enhance Windows, M365 and Teams to take advantage of the local on-device capabilities of Windows 11 PCs including high-quality cameras, sensors, input, and new silicon optimized for AI.

Investments to the Windows platform will enable the Windows ecosystem of developers to effectively reach over a billion customers, to access new capabilities for innovation and differentiation, and reduce development costs. Microsoft is working to build a platform that will expose device capabilities in a scalable manner and deliver AI capabilities in a high-quality and consistent way.

Goals in Windows 11 2022 Update (SV2)

Beginning in the 2022 release of Windows 11 (SV2), new AI capabilities will be added to the Windows platform to enable key hero experiences in the OS. These experiences will leverage the advanced capabilities of neural processing units (NPUs) to drive improved performance and optimal battery life for Windows PCs that have capable hardware, while also delivering higher quality experiences with larger AI models only possible on NPU silicon.

The new built-in camera and audio AI effects available in Windows will improve the collaboration experience for users working and learning from home. Content creation applications for video recording and streaming will also be enhanced by these effects to deliver more compelling content. Speech enabled experiences will deliver more natural and intuitive experiences for accessibility customers and provide productivity benefits for both workers and students. These hardware-enabled experiences meet the needs of key customers and drive purchase intent for Windows devices.

The core investments in Windows 11 SV2 provide the starting foundation of a platform that will enable future AI powered experiences in the OS and capabilities that can be leveraged by 3rd party developers.

Responsible AI

Microsoft believes that the development and deployment of AI must be guided by an ethical framework and has established a company-wide effort around Responsible AI to ensure solutions are secure, private, and compliant with industry regulations. Customers will only embrace the technology if we maintain trust and credibility with our mutual customers. Microsoft believes that Responsible AI is good for the customer and a key strength relative to current in-market AI solutions. Refer to Responsible AI principles from Microsoft to learn more.

NPUs

Neural Processing Units (NPUs) are new types of silicon emerging from the major System on Chip (SoC) vendors that are designed to offload and accelerate Machine Learning (ML) models and deep learning neural networks with optimal performance-to-power efficiency.

NPUs supported in Windows 11 SV2 enable high-fidelity, high-bandwidth processing of advanced ML algorithms to reduce CPU load and minimize impact to battery life, while delivering a higher quality experience than traditional commodity silicon. NPUs are also designed to efficiently process multiple concurrent AI algorithms. Focus will be placed on enabling a class of NPU silicon that can be programmable and accessible by Windows OS experiences and Windows apps. Fixed-function or special-purpose silicon are not supported at this time.

Windows 11 SV2 will focus on the enablement of key hero experiences to establish the platform foundation to fuel future Microsoft AI-powered experiences and ecosystem apps.

For a list of supported SoCs/NPUs and the corresponding feature sets, please work with your Microsoft OEM account representative for this information.

High-level Scenarios

Windows 11 SV2's focus will be on delivering the following hero NPU-experiences:

1) **Windows is more collaborative and productive through built-in Windows Studio Effects:** Windows Studio Effects are new AI-powered camera and audio effects that are built-in into the OS platform and made available to all apps. This will enable customers to communicate, collaborate and stream more efficiently in Real-Time Communication (RTC) apps such as Teams and other 3rd party apps. In addition, built-in camera effects will enhance immersion for content creation applications for capture, streaming, and more. These highly tuned capabilities can also be leveraged by developers to incorporate into their apps and enable new innovative solutions on Windows devices. With a set of advanced ML models directly integrated in Windows that can run across many silicon types at consistent quality, performance, and efficiency, developers can reach a broader set of devices without having invested in the expensive costs of developing using vertically integrated toolchains.

   These built-in effects require a system with a supported NPU due to the bandwidth and latency requirements to deliver a high-quality and performant experience.

2) **Windows empowers users of all-abilities with natural and intuitive voice experiences:** We are bringing features that use state-of-the-art, on-device AI speech models. Windows 11 2022 Update will be the most accessible version of Windows yet. Live captions enable Windows devices to meet the needs of accessibility users who are deaf or hard of hearing. Voice access enables customers to use their voice to control their PC and interact with applications. These experiences offer increased productivity and comprehension for workers and students, while being more natural and intuitive to use.

   These experiences will be available to all Windows PCs but will run better if the system includes a supported NPU.

System Design Considerations

Supported NPU Silicon

Windows 11 SV2 will support a select set of SoCs and NPUs from major IHV silicon providers. The experience feature set enabled for each NPU will vary based on the total compute capacity (TOPs) and performance level (TOPs/watt) of the silicon.

For a list of supported SoCs/NPUs and the corresponding feature sets, please work with your Microsoft OEM account representative for this information.

Real-Time Communications (RTC) Optimized Hardware Guidance

Windows 11 SV2 AI-experiences will leverage on-device capabilities of the system for input, such as high-quality cameras, microphones, and speakers. The RTC Optimized Hardware advocacy is a consolidated set of hardware requirements to guide OEMs to build Windows PCs with the right cameras, mics, and speakers that will enable high-quality video and audio quality for a wide range of experiences on Windows 11.

Refer to the RTC Optimized Hardware advocacy and validation toolchain on Microsoft Collaborate and design your systems accordingly based on price-band. The latest specification incorporates incremental additions to support SV2 NPU-enabled experiences. Additional HLK and test tools will be used for scenario specific validation in later sections of this document.

Azure Speech: Edge Devices Specification

The Azure Speech: Edge Devices specification provides OEM device makers with best practices for device design and a validation framework to build speech-enabled devices. The Azure Speech device hardware specifications for Windows PCs are reflected in the RTC Optimized Hardware Guidance. We recommend that you review the specification and use the tool chain to optimize your devices for the best speech experience on Windows.

Additional SoC/NPU design considerations

Depending on the SoC/NPU in the system, additional design considerations may apply. These will be documented separately in IHV-specific documentation as appropriate. Please ensure that you coordinate with your silicon IHV for additional design guidance.

Why Windows Studio Effects?

Windows Studio Effects aims to deliver customers and app developers with a consistent set of high-quality effects, powered by advanced AI models, while delivering enhanced performance and battery life across a diverse set of hardware.

Desktop AI effects have traditionally been a difficult space. The vast ecosystem of devices with varying specs makes it challenging for applications to leverage hardware acceleration for inferencing, leading to AI-powered effects often running entirely on less efficient general purpose CPU cores. In addition to lower efficiency in a world where battery life is more important than ever, the need for AI-powered effects to work on devices at all price points can lead developers to reduce the complexity/fidelity of their effects to be able to run on low-cost CPUs. This leads to an experience that compromises quality, battery life and thermals. Additionally, AI effects provided by apps are scoped to the app itself – a customer can turn on Background Blur in Microsoft Teams, but not in a web app, creating an inconsistent experience.

For camera AI effects, OEMs, IHVs and camera manufacturers have attempted to address these issues outlined above by building proprietary AI effects directly into their cameras. Direct integration into the camera allows the *device* to offer effects for any app to leverage instead of relying on an *app* to offer its own built-in effects. Some manufacturers may choose to optimize their effects for a specific CPU, use GPU acceleration, or embed an NPU directly into the camera. However, there are two key issues that ISVs encounter when individual cameras use these solutions:

1. **No standard mechanism for an app to discover, control, and coordinate camera effects.** There is no standard way for applications to understand and control the effects that a camera can provide. This can create confusion for mainstream customers due to the layering of effects. The following challenges can occur for a camera that directly includes its own Background Blur effect:
    a. The camera's Blur could be turned on, but Microsoft Teams reports Blur is off because it's unaware of it
    b. The camera might support a better Blur (quality and/or power), but the customer does not get to take advantage of it when they turn Blur on in Microsoft Teams
    c. The camera might support a lower quality or lower performance Blur, and the customer may use it without being aware that there is a higher fidelity and more performant option

2. **Inconsistent quality of effects.** When an application provides an in-app control for an effect (e.g. Background Blur), they are taking accountability for the *feature* as part of their product. This includes taking responsibility for the customer satisfaction and technical support costs associated with exposing an implementation that is potentially poor or unpredictable in quality.

    If an application wishes to provide a Blur control, and its options include:
    a. Provide a consistent in-app implementation that is controlled end-to-end
    b. Implicitly trust that the implementation provided by any camera meets their quality and performance criteria.

    Developers are realistically choosing option (a) in most cases.

Microsoft Effect Pack aims to solve these problems in two key ways:

1. **Standardized Control Interfaces (Camera DDIs and APIs)** in the Windows OS for each of the effects, that any application can use to discover if effects are supported by the camera, turn them on/off, and access any metadata they produce.
2. **Consistent AI Models** provided by Microsoft and compiled/optimized for all supported NPUs. This means that once the MEP effects are tested and validated for quality, an application can trust that the MEP effects will produce consistent results for the same input stream, even across devices/silicon.

Windows Studio Effects Deployment Model

Windows Studio Effects are delivered as a driver package (interchangeably referred to as the MEP driver package) that contains core components necessary to enable NPU-experiences, including camera and audio AI effects.

The MEP driver package contains NPU-dependent binaries and tuned AI algorithms. Distinct MEP packages are provided for each NPU-type and each package will only contain the necessary components for a given system/SoC.

Configuration and deployment of the MEP driver package requires the following steps:

1) **OEM opts in the NPU-system for Windows Studio Effects to enable deployment targeting.** Opt-in enables the binding mechanism to deploy the MEP driver package to the NPU-system through OEM pre-installation or Windows Update.
2) **OEM opts in front-facing integrated camera AI effects.** Camera AI effects requires explicit opt-in as the OEM needs to specify the correct front-facing camera to use the MEP effects. This allows the platform to associate the effects to the appropriate camera in the case that there are multiple integrated front-facing cameras on the system. Refer to **'Enabling Camera AI Effects'**.
3) **Preinstall Microsoft Effect Pack package.** For the best out-of-box customer experience and to enable MEP effects to be available by default on a new NPU-system, preload of the MEP driver package is recommended.
4) **(Optional) Deploy Microsoft Effect Package via Windows Update.** Opting in MEP at the system level in step #1 enables MEP to be deployable through Windows Update. The steps for in-market update varies based on SoC/NPU platform.

Opt-in method varies by IHV silicon. Refer to IHV specific documentation for more details on how to opt-in the Windows Studio Effects for an NPU-based system.

Enabling Windows Studio Camera Effects

Windows Studio Effects camera effects are system-wide effects that are made available to all apps running on an NPU-enabled system. The effects are optimized for integrated front-facing cameras. OEMs are required to opt-in the front-facing camera of the system to enable these effects on an NPU-enabled system.

The following camera AI effects in MEP include:

- **Background Blur:** Background Blur conceals the user's background to minimize distractions and increase privacy, while keeping the user clear and in focus.
- **Portrait Blur:** Portrait Blur is a stylistic version of Background Blur that softens the background to give an aesthetic effect that is common in portrait photography.
- **Eye Contact "Standard" v1:** Eye Contact adjusts the user's gaze to give the impression that the user is making direct eye contact with the viewer of the video to increase engagement and personal connection. Eye Contact v1 supports landscape orientation devices only.
- **Eye Contact "Standard" v2:** Eye Contact v2 is a higher quality version of v1 and provides support for multiple device orientations including portrait mode. Eye Contact v2 requires a more performant NPU than v1 to enable these additional capabilities.
- **Eye Contact "Enhanced":** Eye Contact Enhanced is a more advanced version of the Eye Contact "Standard" effect that adjusts the user's eyes for scenarios when they are looking down and their eyes are moving side to side to read a script or document during a presentation. (Note that Eye Contact Enhanced has been postponed for the future release of Windows. Timing has not yet been finalized.)

- **Automatic Framing (no Super Resolution):** Automatic Framing automatically keeps the user centered and in focus in the camera frame.
- **Automatic Framing (with Super Resolution):** Automatic Framing automatically keeps the user centered and in focus in the camera frame. This version of Automatic Framing uses super-resolution to greatly enhance the visual quality of the user.

Feature support varies by NPU. For a list of supported features by NPU, please refer to NPU-vendor specific documentation or work with your Microsoft OEM account representative for this information.

Microsoft recommends OEMs to use the built-in effects from the Microsoft Effect Pack on NPU-supported systems and to avoid shipping duplicate effect implementations. *Refer to "Co-existence of MEP effects and 3rd party OEM/IHV effects", for further details.*

Native Settings Integration for Camera Effects

The camera effects will have native integration into Windows Settings for users to toggle them on or off system wide.

Note: Support for camera effects will vary based on system/NPU silicon in the device. Settings toggles for individual effects will be hidden if not supported on the device.

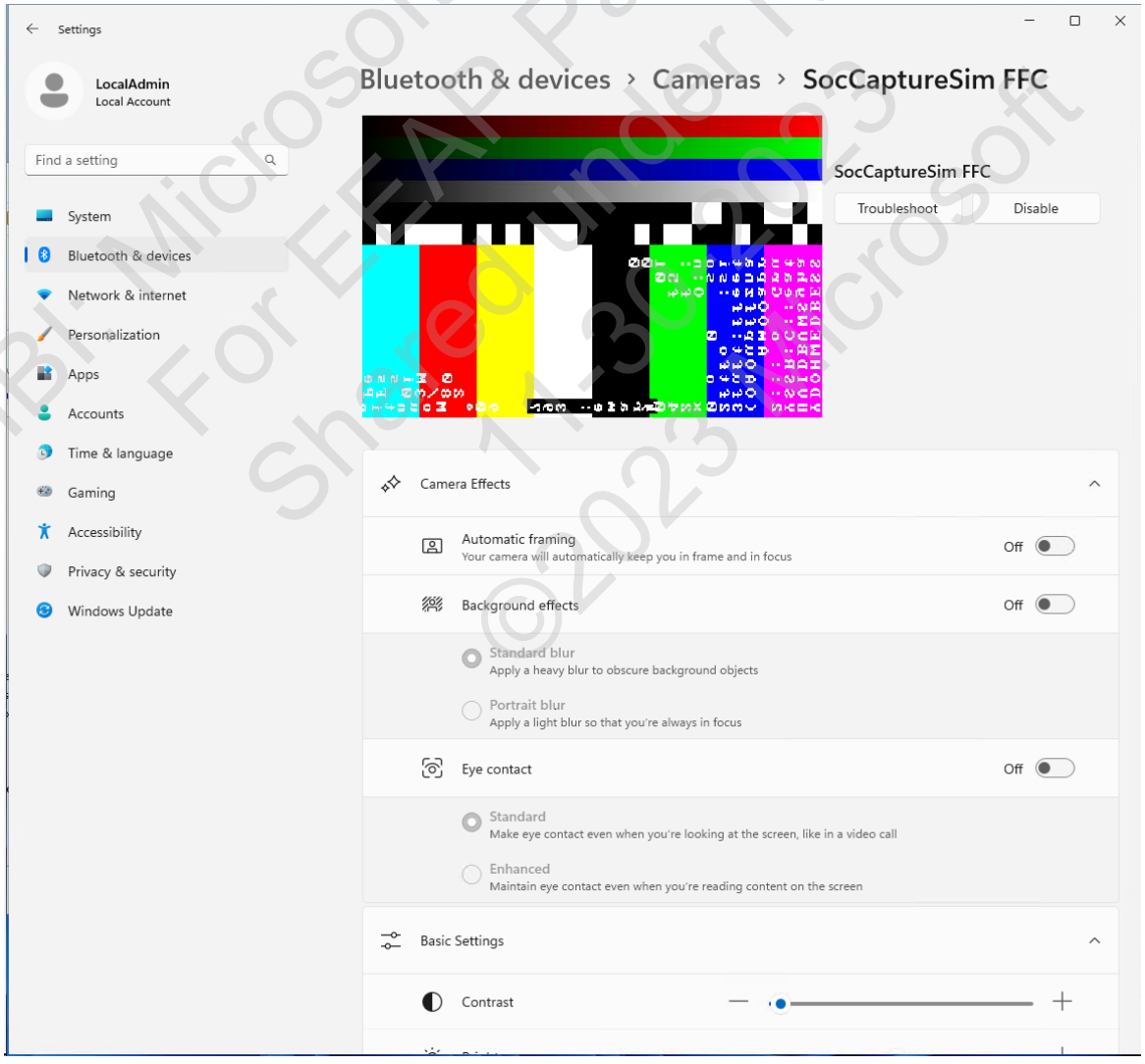


Figure 1 - Windows Settings Camera Page (Concept UX with Microsoft built-in Camera effects)

By default, camera AI effects settings will be turned off for the user. When the user turns an effect on, the user's preferred value is stored by the camera pipeline and applied to the camera's current value every time an app starts the camera. For more information on the handling of default settings by the OS and applications, refer to *'How default values for Camera Settings work'*.

Windows Studio Effects on taskbar Quick Settings

With the release of the Windows 2023 February CI release, Windows Studio Effects can now be accessed directly from Quick Settings on the taskbar. This makes it quick and easy to enable and configure camera effects (Background Blur, Eye Contact, and Automatic Framing) and audio effects (Voice Focus).

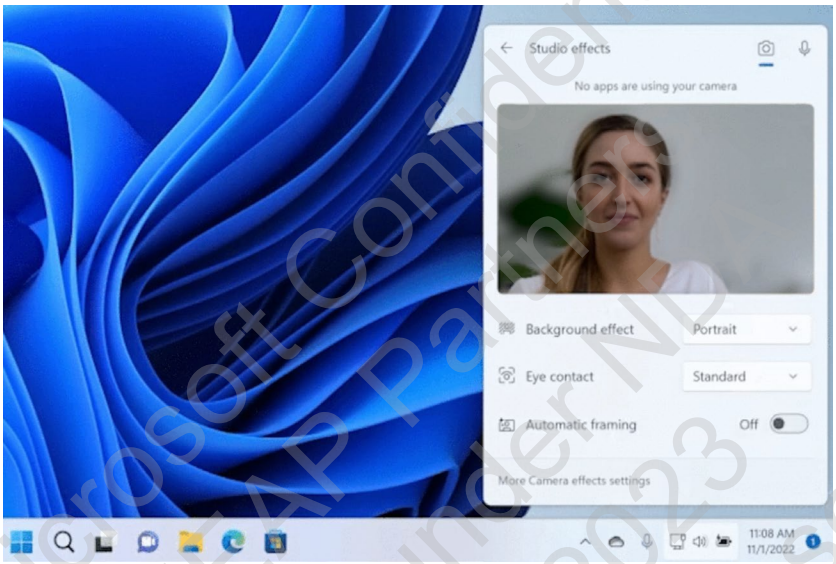


Figure 2 Windows Studio Effects in Windows taskbar Quick Settings

Users can still access these effects in the Settings pages if desired. This menu will only be available on devices with a supported NPU and if Windows Studio Effects are opt-ed in by the OEM.

OEM Opt-in of MEP Camera AI effects

OEM opt-in of the front-facing integrated camera is required by the OEM or IHV owning the camera driver to enable the built-in effects on the system. Integrated front-facing USB and MIPI-CSI cameras are supported for opt-in. When an integrated front-facing camera is opted in, Windows will enable the full set of effects that the camera and NPU are capable of handling. The current opt-in model for the camera effects is an 'all or none' mechanism. Opt-in or opt-out of individual effects is not supported.

For a list of supported SoCs/NPUs and the corresponding feature sets, please refer to NPU-vendor specific documentation or work with your Microsoft OEM account representative for this information.

Key System and Camera points to note for MEP Opt-In

Key changes made in this section in v1.1 documentation release are highlighted in blue for clarity purposes

1. The camera to be opted-in must be USB or MIPI based built-in device with an AVstream-based driver
    a. At least one video record pin or video preview pin must be present.
    b. At least one video MediaType on the record stream that is compatible. (Limited to MJPG, YUY2, NV12)
    c. At least one MediaType on the record stream that passes the following constraints.

2. NPU is required for camera AI effects of Microsoft Effect Pack to work. Depending on the variety of camera modules, diversity of platform configurations, and driver deployment infrastructure, OEMs can choose one of the following MEP camera Opt-in options to meet their design, engineering support, and logistics management processes:
    a. **NPU-targeted Opt-in**: Use the camera driver INF method to enable MEP Opt-in only onto systems with NPU. This can be helpful in targeting a specific SoC that may have variants with and without NPU. OEMs can use CHID targeting to deploy the Camera driver INF onto certain systems.
    b. **Camera module-based Opt-in to target multiple designs sharing the same camera module:** Use camera firmware to enable MEP Opt-in onto all systems, and only disable MEP Opt-in on systems without NPU by using camera driver INF **AddReg** directive to 0 to overwrite the registry key. Properties defined within a camera driver INF take precedence over properties defined in camera firmware. This option gives OEMs the flexibility to efficiently enable the Camera opt-in on a large number of designs that share a common camera module and opt-out MEP on select systems using the Camera driver INF.
    c. **Separate Camera module-based Opt-in:** Have two variants of a camera module with separate DPN, one with firmware to enable MEP Opt-in, and the other with firmware without MEP Opt-in.
    d. **Separate Camera driver-based Opt-in:** Have two camera driver INFs, and all systems should apply to either one of them. Apply the camera INF specifying MEP opt-in on systems with an NPU. Apply the camera INF without MEP opt-on on systems without NPU.

    Refer to detailed guidance in **OEM/IHV Opt-in Mechanism of MEP on the front-facing camera** section.

3. Windows Studio Camera affects are applied to resolutions less than or equal to 2560x1440 and greater than or equal to 480x360.
    a. Applications may choose to use the camera at higher resolutions if the hardware permits, but AI effects will not be applied.
    b. It is expected that greater than 2560x1440 resolution will not be exposed to the MediaType list for video recording in the Camera App, and for most other usages in other apps that are not profile-aware (Discover and select camera capabilities with camera profiles – UWP applications | Microsoft Learn). MEP has a throughput limit; currently can only process with effects a stream at up to 1440p at 30fps.
    c. That said, for photo capture in the Camera App, the higher resolutions (i.e., 4k) should be usable, but effects will not be applied (i.e. there are no in-app toggles to set Background Blur/Automatic Framing, etc. in the Windows Camera App in photo capture mode, and if using settings page to enable the effects then they will not apply to the photo capture mode). This can be seen as a "passthrough" mode of operation.
    d. Similar to the photo capture "passthrough" mode, Windows Studio Effects also supports the video recording passthrough at higher resolution and framerate in MEP versions 1.0.25 or greater. Enabling passthrough support in the Windows Camera App is under in a future release (timing is TBD).

4. There is no plan in the near future to apply effects at higher resolutions.
5. Frame Rate equals or greater than 15 and less than or equal to 30 frames per second.
6. The effects will be applied only when the Microsoft Effect Pack for camera is installed on the system.
7. The camera and system must pass HLK, both before adding the opt-in mechanism and after opt-in.
8. The camera must be front (user facing) camera. The camera location information in PLD is recommended for the best experience. MEP is not supported on world facing, movable, convertible, or external cameras.
9. At most only one camera on the system can opt-in
10. All profiles exposed by the camera driver will still be present after opt-in, but the following ones supporting Windows Studio effects will be altered to correctly express the set of output MediaTypes when the camera is opted in:
    a. KSCAMERAPROFILE_Legacy
    b. KSCAMERAPROFILE_VideoRecording
    c. KSCAMERAPROFILE_VideoConferencing
11. If present, the Face Authentication camera profile (KSCAMERAPROFILE_FaceAuth_Mode), will be available only to Windows Hello and will not subject the frames to any Windows Studio effects.
12. An orientation sensor is recommended for devices that have multiple postures, including tablets, 2-in-1 convertibles, and 2-in-1 detachables. This is to identify portrait versus landscape camera stream captures.
    a. If the orientation hardware is not present in the system, landscape camera orientation will be used by default.

Post Opt-in Expectations

Once a camera is opted into the Microsoft Effect Pack, the camera output that is offered to consuming applications gets altered (constricted), whether the application leverages any of these effects or not, as listed here:

1. When an app opens the camera without a specific profile, only one video pin (record pin if possible, else the preview pin) will be exposed to the app. Other pins/streams from the camera will get dropped.
2. The single output pins available to the application will have resolutions, frame rates and MediaTypes that are supported by MEP
    I. Video resolutions less than or equal to 2560x1440 and greater than or equal to 480x360 and frame rate equal or greater than 15fps and less than or equal to 30fps.
    II. All output MediaTypes will be of NV12 subtype equivalents to the qualified input MediaTypes as defined in (1)(b) in the **'Key System and Camera points to note for MEP Opt-In'** section.
3. Photo pin limitations – Independent Photo pin will not be made available to the application. Consequently, Photo resolution will be limited up to the Preview/Record video resolution. This is similar to a single pin USB camera without Method 2 or Method 3 photo support.

    Note: The following functionality has been made available in the Windows 11 SV2 (22H2) '8D' servicing release at the end of August, 2022:

    When a profile-aware app opens the camera explicitly using any other profile than the ones with which Windows Studio camera effects are supported (KSCAMERAPROFILE_Legacy, KSCAMERAPROFILE_VideoRecording, KSCAMERAPROFILE_VideoConferencing) such as the 'Photo' profile (KSCAMERAPROFILE_HighQualityPhoto):

1. *Other pins than the one leveraged by Windows Studio are exposed (such as a dedicated photo pin, a separate preview pin) to use and all MediaTypes from these pins are available.*
2. *The MediaTypes exposed out of the pin leveraged for Windows Studio effects are unconstrained.Windows Studio camera effects shall not be supported.*

This allows profile-aware apps, such as the Windows Camera app when engaging the photo capture mode, to be able to access the photo and preview pins at the full camera resolution supported albeit without the ability to toggle any Windows Studio effects.

4. *For more information in declaring KSCAMERAPROFILE_HighQualityPhoto, refer to the following guidance:* Camera Profiles - Windows drivers | Microsoft LearnNumber of Cameras opting in is limited to 1, due to computational limitations on NPU.
5. The non-RGB Sensor pins will not be available to apps and will not have the effects enabled. If an MEP opted-in camera is used for Windows Hello, no MEP effects are applied to the camera stream.

   Note: The following functionality has been made available in the Windows 11 SV2 (22H2) '8D' servicing release:

   > *If a consuming application sets a profile other than KSCAMERAPROFILE_VideoRecording, KSCAMERAPROFILE_VideoConferencing, and KSCAMERAPROFILE_Legacy, then it can access the non-RGB sensor pins.*

6. No compressed media types (including MJPEG and other temporally compressed) will be made available to the apps.
7. No custom metadata payload to carry across - metadata might be invalidated by the signal processing here
8. If the system includes 3rd party OEM/IHV implementations of camera effects occupying the same underlying DDI's used by effects included in MEP, the Microsoft built-in effects will override and take precedence as the default implementation for the end customer. For more information on the DDIs used by MEP effects, refer to the section **"Co-existence of MEP effects and 3rd party OEM/IHV effects".**
9. The following DDIs are always unsupported:
   I. KSPROPERTY_CAMERACONTROL_EXTENDED_PHOTOMODE
   II. KSPROPERTY_CAMERACONTROL_EXTENDED_WARMSTART

OEM/IHV Opt-in Mechanism of MEP on the front-facing camera

The camera effects are available on a front-facing camera only if the OEM wishes to opt-in using a registry key (regkey).

The opt-in regkey under the Device Interface Node registry is:

FSMEnableMsEffects: REG_DWORD: 0x1

Note: Some camera drivers may optionally expose legacy device interfaces apart from KSCATEGORY_VIDEO_CAMERA, including KSCATEGORY_CAPTURE and/or KSCATEGORY_VIDEO.

In these cases, the OEM/IHV will also need to apply the FSMEnableMSEffects opt-in regkey to these device interfaces. This will ensure that these interfaces can be properly enumerated by the OS and applications with MEP effects are correctly registered.

This registry entry may be added by any of these supported methods:

1. **Custom INF:** Using a custom INF by adding the registry entry through the **AddInterface** directive.
    a. For more information on custom INF files, refer to Using an Extension INF File and INF AddInterface directive
    b. Note: If the system is using the inbox UVC class driver for an internally connected USB camera, a custom INF can be used to add the registry entry if 'usbvideo.inf' is referred to load the inbox UVC class driver. Refer to Providing a UVC INF File for more information.
2. **USB Video Camera (UVC) MS OS descriptors** – MS OS descriptor information is written to the **Device Node** rather than the **Device Interface Node**.
    a. The registry entry name must be prefixed with **"UVC-"**. This informs the inbox UVC class driver to migrate this registry entry (minus the prefix) to the Device Interface Node.
    E.g., UVC-FSMEnableMsEffects: REG_DWORD: 0x1
    b. There must be only one camera interface node under the device node, else all color cameras under the Device Node will get opted-in.
    c. For more information on MS OS descriptors, refer to Microsoft OS descriptors and Create device property keys from the MS OS descriptor in USB Video Class (UVC) camera firmware

Note: Opting-in into camera effects is orthogonal to opting-in to OS supplied Face Detection via Platform DMFT chaining. The INF entries are different, and either one or both could be opted-in.

Specifying camera physical location information (PLD)

The camera effects can be optimized if the relative location of the integrated camera on the panel is known by populating the ACPI PLD information. For example, the quality and accuracy of Eye Contact is optimized based on the understanding of the precise camera position.

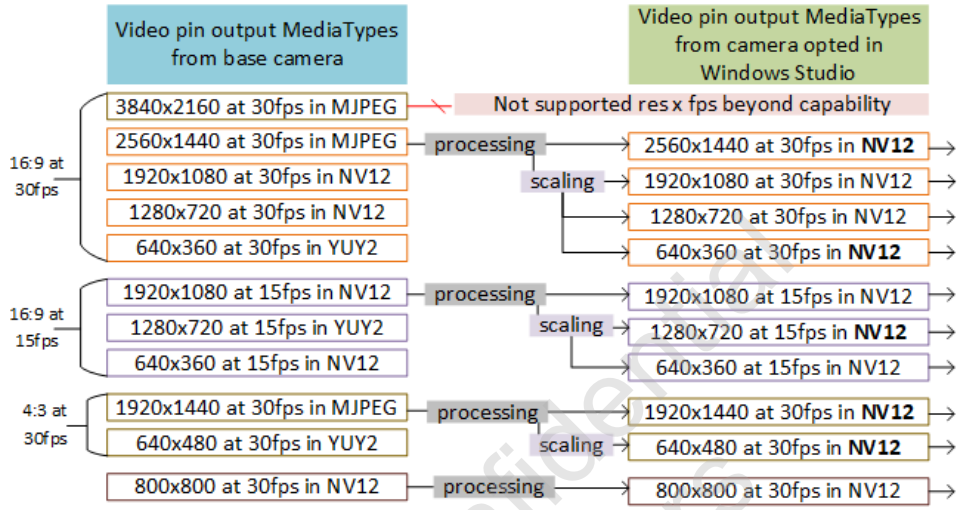
The necessary _ACPI_PLD_V2_BUFFER members to be populated include:

- Panel (Mandatory)
- VerticalOffset (Optional, Recommended)
- HorizontalOffset (Optional, Recommended)

Refer to this documentation link for more details: Driver support for camera orientation - Windows drivers | Microsoft Docs

Leveraging higher resolution processing to improve visual quality

To enhance image quality when applying effects such as "Automatic Framing", Windows Studio can be set to only leverage and process the highest resolution supported by MEP exposed by the camera driver and scale accordingly to the target resolution at all times. If the app's desired resolution triggers a higher resolution from the source to be used under the hood by MEP as input, it may consume more compute resources and therefore come as a tradeoff for image quality.. The highest resolution chosen correlates with the desired framerate and aspect ratio of the MediaType requested by the application at up to the maximum input capability in MEP i.e. up to 1440p (see below diagram)



Some notes about

To enable this option the following DEVPROPKEY must be set (refer to INF AddProperty directive - Windows drivers | Microsoft Learn):

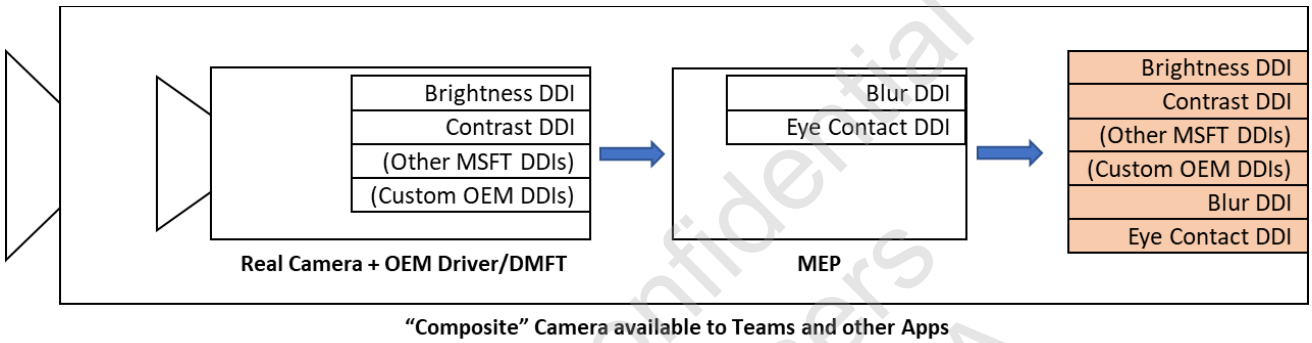
- Example of an .inf excerpt for enabling the high resolution devpropkey:

```
[OPTIN_CAMERA_INF.Interfaces]
AddInterface=%KSCATEGORY_CAPTURE%,Global,OPTIN_CAMERA_INF.Interface
AddInterface=%KSCATEGORY_VIDEO_CAMERA%,Global,OPTIN_CAMERA_INF.Interface
AddInterface=%KSCATEGORY_VIDEO%,Global,OPTIN_CAMERA_INF.Interface

[OPTIN_CAMERA_INF.Interface]
AddReg=OPTIN_CAMERA_INF.Interface.AddReg
AddProperty=Camera.AddProperty

[OPTIN_CAMERA_INF.Interface.AddReg]
; Add reg key to opt into Windows Studio
HKR,,FSMEnableMsEffects,0x00010001,1

[Camera.AddProperty]
; Set property to operate Windows Studio in high resolution pipeline mode
{AA3E8B1E-B590-4E50-90C6-780AEC4EB4D9},2,7,,1
```

Co-existence of MEP effects and 3rd party OEM/IHV effects

MEP effects are chained after the OEM Driver/DMFT. The OEM's camera features and supported DDIs are supplemented by MEP's effects to form a "Composite" camera that is made available to applications.

The following figure represents how MEP is chained in the camera pipeline and made available as a "Composite" camera to consuming applications:



When Microsoft Effect Pack effects are opted in on an NPU-supported system, the following effects will occupy these DDIs:

MEP Camera effects	**DDI**
Background Blur (Standard & Portrait)	KSPROPERTY_CAMERACONTROL_EXTENDED_BACKGROUNDSEGMENTATION
Eye Contact (Standard & Enhanced)	KSPROPERTY_CAMERACONTROL_EXTENDED_EYEGAZECORRECTION
Automatic Framing	KSPROPERTY_CAMERACONTROL_EXTENDED_DIGITALWINDOW and KSPROPERTY_CAMERACONTROL_EXTENDED_DIGITALWINDOW_CONFIGCAPS

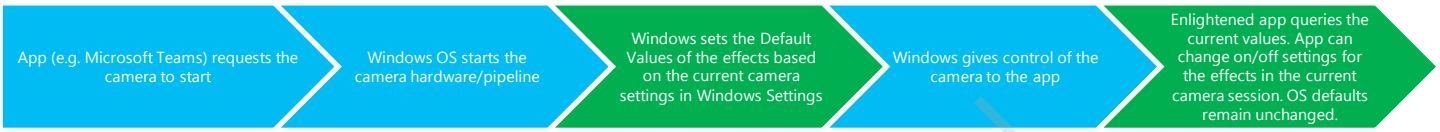
If an OEM has opted into MEP on an NPU-supported system and there exists an OEM/IHV implementation of an effect that uses one of the same underlying DDIs shown above, this results in the existence of duplicate effects on the system. In this scenario, the Microsoft built-in effects will override and become the default set of effects that the end user will control from within the Windows Settings Camera page. The MEP effects are always applied last at the end of the driver chain and the platform will guarantee that duplicate implementations of the same effects for a given DDI will not conflict with MEP effects.

How default values for Camera Settings work

Default values for camera settings are stored on a per-camera, per-user, and per-machine basis. When any app opens the camera, Windows initializes the current value of the controls to match the default values for the current user in the following sequence:

A customer can use their camera in another application and open the Windows Camera Settings page at the same time. Within Windows Camera Settings, any changes to the effect toggles will update the default value. This also changes the current value concurrently, impacting the active stream going into the running application that is using the camera. A message is displayed when the Windows Camera Settings Page is opened while another app is using the camera to indicate that changes made will impact the active camera stream.

Camera application developers can enlighten their apps by using the camera DDIs to query for the built-in MEP effects and toggle them individually, giving them choice and control over the effects within their app experience. The following sequence demonstrates how an enlightened application can supersede the default Windows Camera Settings:



OEMs may also develop Companion Apps for their cameras that can replicate and/or extend the functionality of the Windows Camera Settings Page.

Refer to Cameras, Sensors and Buses Windows 2022 Experiences and Windows Studio Overview & OEM Extensibility for Cameras for more details on how companion apps can synchronize settings seamlessly with the Windows Camera Settings page and offer the ability to extend with manufacturer differentiation, and how apps can launch the Windows Camera Settings Page.

More details on how ISV developers can use Windows Studio Effects can be found in the following resources:

Windows Studio Overview & ISV Integration Guidance
Windows Studio Overview & ISV Integration Guidance (TLDR Version)

Design considerations for OEM effects extensibility and co-existence

Microsoft recommends OEMs to use the built-in effects from the Microsoft Effect Pack on NPU-supported systems and to avoid shipping duplicate effect implementations.

If OEMs/IHVs wish to implement additional differentiated effects, Microsoft recommends that the OEM chooses effects that do not overlap with the built-in Microsoft effects. OEM/IHV-provided effects that perform heavy computation on CPU or GPU are strongly discouraged, as they result in poor system performance and battery life without the impact of each effect being clear to the customer.

OEM/IHVs who want to add differentiated camera effects tied to a physical camera must be implemented using DMFTs. In Windows 11 SV2, the number of allowable chained DMFTs is increased from 2 to 4 to increase flexibility for hardware vendors.

For more guidance on camera effects extensibility and DMFTs, refer to the following resources:

- Device MFT design guide
- Configuring the DMFT chain order
- Cameras, Sensors and Buses Windows 2022 Features Update
- Cameras, Sensors and Buses Windows 2022 Experiences
- Cameras, Sensors and Buses Windows 2023 Experiences
- Windows Studio Overview & OEM Extensibility for Cameras

Virtual Camera limitations with MEP

It is not recommended for OEMs to implement additional effects using Virtual Cameras (SCS Virtual Camera, Custom Media Source, DirectShow Sources, or custom AVStream drivers) as they can result in limitations and incompatibilities:

1) A Virtual Camera is represented as a separate camera device within Windows Settings and Windows applications. Users are required to explicitly select this device as a source in a consuming application, such as the Windows Camera app, Teams, and other applications that use the camera stream. This enables customers to have full explicit control to select a virtual camera and image processing effects associated to the software camera.
2) Due to technical limitations in how Virtual Cameras are architected in the platform, Microsoft built-in system effects delivered through the 'Microsoft Effects Pack' is not supported on a Virtual Camera device.
3) Microsoft built-in camera effects can only be applied to physical integrated front-facing cameras and require the system to pass HLK testing to ensure proper system compatibility.
4) Microsoft Effects Pack must always be the last "block" in the chain, so that when it implements an effect and an enlightened application wishes to use the effects in Microsoft Effects Pack, the application can discover and explicitly control the Microsoft effects.
5) Physical cameras in a system are not allowed to be hidden by the manufacturer in favor of a virtual camera device that wraps the camera. This creates a confusing experience for the customer where:
    a. There are two cameras available to applications that behave differently from each other, and customers must understand which camera to select and the concept of virtual cameras
    b. As described above, the Microsoft Effects Pack experience only works correctly with enlightened applications if the customer selects the "real" camera that is opted into Microsoft Effects Pack. If the customer selects the "virtual" camera, an enlightened application is not able to discover and control the Microsoft Effects Pack experience on that camera
6) Introducing a virtual camera adds another layer of compute complexity, reducing battery life and system performance.
7) Virtual Cameras based on DirectShow Source technology are only available for applications using the DirectShow APIs. Applications using modern Media Foundation (MF) based APIs are unable to discover and use DirectShow Sources.

Validating Camera Effects

New SV2 HLK requirements have been added to ensure proper configuration of the built-in camera effects and ensure systems meet WHCP.

Device.Streaming.Camera.Sharing.CameraAIEffects

New tests include:

- Camera Driver System Test - Mediacapture - TestWindowsCameraEffectStreaming
- Camera Driver System Test - Mediacapture - TestCreateWindowsCameraEffectMediaSource
- Camera Driver System Test - Mediacapture – TestEffectOptInCameraEnumeration

The camera and system must pass HLK before adding the opt-in mechanism and after opt-in. Running the HLK tests before MEP is opted-in ensures there are no regressions. Only the results package after MEP is opted-in is required to meet WHCP.

Enabling Voice Focus (Deep Noise Suppression)

Windows 11 SV2 will introduce a new built-in microphone AI effect that is made available system-wide to all apps on NPU-enabled systems. This effect, Voice Focus, is optimized and available for integrated microphones.

Voice Focus uses an advanced deep noise suppression algorithm to eliminate a variety of background noises and enhances the clarity of the speaker's voice. This results in more efficient and productive conversations in video conferencing calls and improved audio in recording applications that leverage the audio stream.

Voice Focus has SoC/DSP-specific design considerations and requires additional validation processes.

Support for Voice Focus varies based on NPU. Refer to IHV specific documentation or work with your Microsoft OEM account representative for this information.

Microsoft recommends OEMs to use the built-in audio effects from the Microsoft Effect Pack on NPU-supported systems and to avoid shipping duplicate effect implementations.

Refer to ***"Co-existence of MEP audio effects and 3rd party OEM/IHV audio effects",*** *for further details.*

Native Settings Integration for Audio Effects

Building from what exists today, AI Audio Effects will be available in the System Sound Properties page associated with a specific device. This will provide users with a dropdown selector to pick enhancements available on their device. When Voice Focus is selected, users will have the option to toggle Voice Focus on or off.

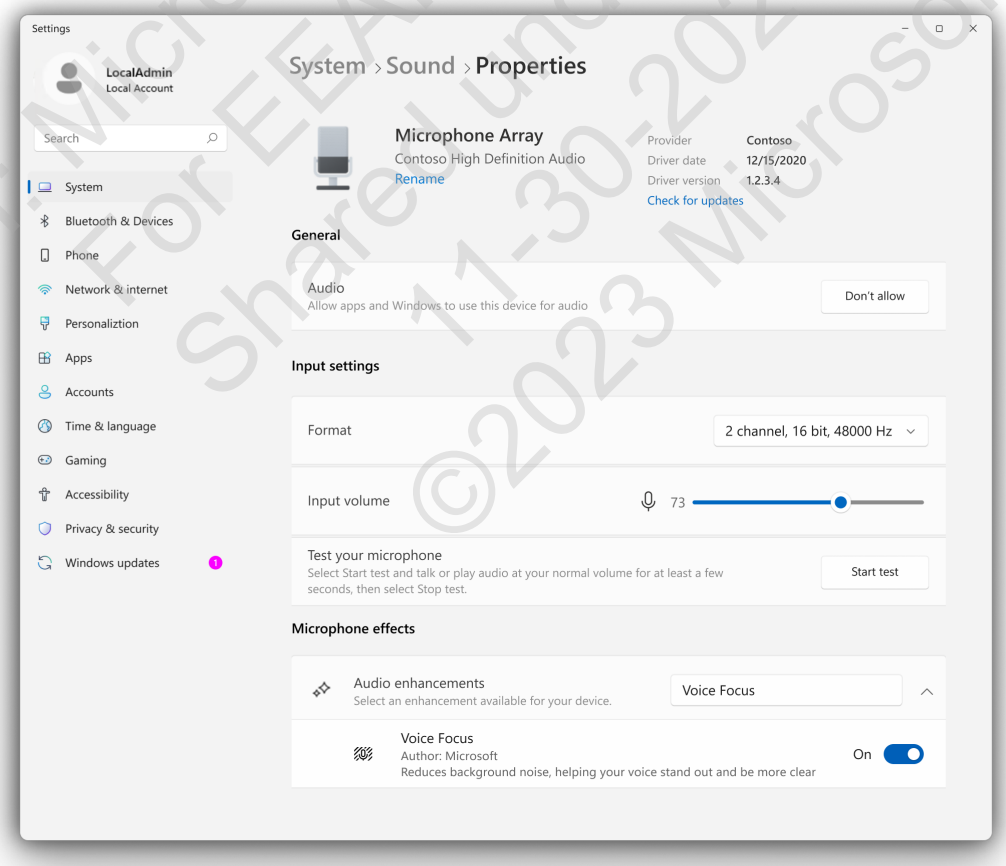


Figure 4 - Windows Settings Sound Properties Page (Concept UX with Microsoft built-in Audio effects)

Devices that ship with Voice Focus will automatically have Voice Focus available for use. In other words, users will not have to select Voice Focus from the dropdown in Figure 4, it will already be selected. By default, audio AI effects settings will be turned off for the user. When the user turns an effect on, the effect is applied to the audio pipeline and made available to all apps that consume the pipeline.

Audio application developers can use audio APIs to query for the built-in effects and toggle them individually, giving them choice and control over the effects within their app experience. These will be available through the same set of APIs APO developers can expose to applications today, available here:

https://docs.microsoft.com/en-us/windows-hardware/drivers/audio/windows-11-apis-for-audio-processing-objects

Audio Pipeline Hardware Variation based on SoC Platform

Depending on DSP hardware, the audio effect package will differ. If a device has hardware DSP available that provides Acoustic Echo Cancellation (AEC), the audio effect package will contain solely Voice Focus AI-based noise suppression. If a device does not have hardware DSP that provides AEC, the audio effect package will apply Microsoft's Voice Clarity AEC before applying Voice Focus AI-based noise suppression. Refer to NPU-vendor specific enablement documentation for more details

Audio Processing Modes

Voice Focus will only be applicable to the **Default** and **Communications** processing modes. For more information on the full set of processing modes available, you can learn more at the following link: https://docs.microsoft.com/en-us/windows-hardware/drivers/audio/audio-signal-processing-modes

Voice Focus will not apply to the **Speech** or **Raw** processing modes. As it exists today, **Raw** mode specifies that there should not be any signal processing applied to the stream. An application can request a raw stream that is completely untouched and perform its own signal processing. Voice Focus will not apply to **Speech** mode in favor of the inbox speech APO dedicated to speech processing.

Co-existence of MEP audio effects and 3rd party OEM/IHV audio effects

Devices that are currently in-market and provide their own APOs will have an updated UI available in the same Sound Properties device page that allows for a dropdown selection between “on” and “off.” See Figure 5.

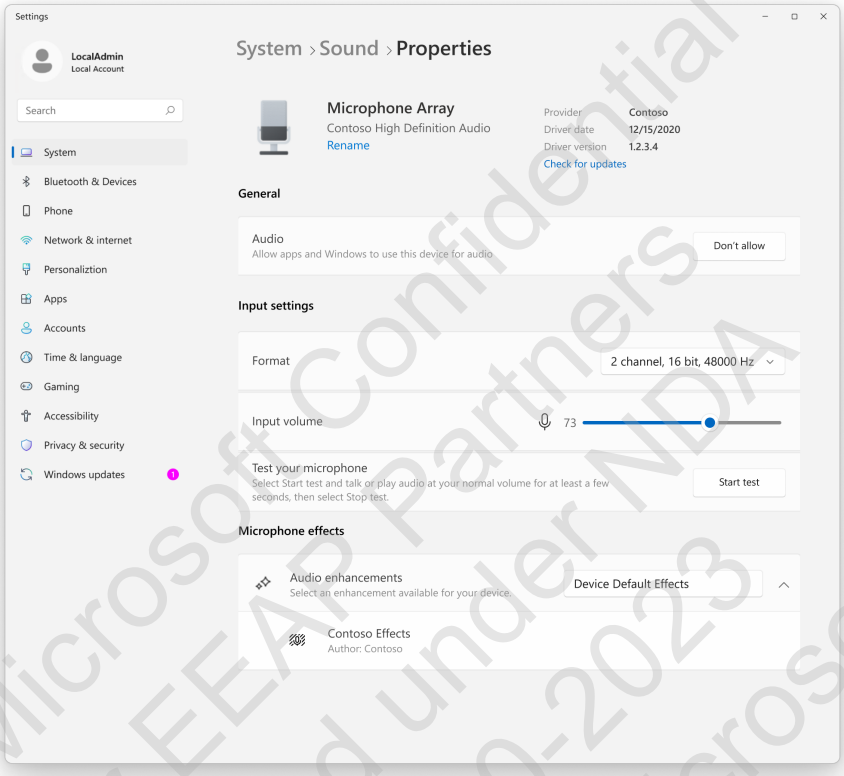


Figure 5 - Windows Settings Sound Properties Page (Concept UX with 3rd Party APOs)

Microsoft recommends OEMs to use the built-in effects from the Microsoft Effect Pack on NPU-supported systems and to avoid shipping duplicate effect implementations.

If OEMs/IHVs wish to implement additional differentiated effects, Microsoft recommends that the OEM chooses effects that do not overlap with the built-in Microsoft effects. OEM/IHV-provided effects that perform heavy computation on CPU or GPU are strongly discouraged, as they result in poor system performance and battery life without the impact of each effect being clear to the customer.

Windows Studio Effects Experience and Scenario Validation

We recommend partners follow the Windows 11 SV2 - Windows Studio Experience & Scenario Test Guide document for an understanding of the supported capabilities, behaviors, and constraints of the experience of the Windows Studio Effects. This document also contains recommended test cases to verify successful deployment and functionality that are recommended to OEMs and ODMs to test. We encourage partners to apply their own testing methodologies in addition to our recommended set of tests to expand validation coverage of the experience.

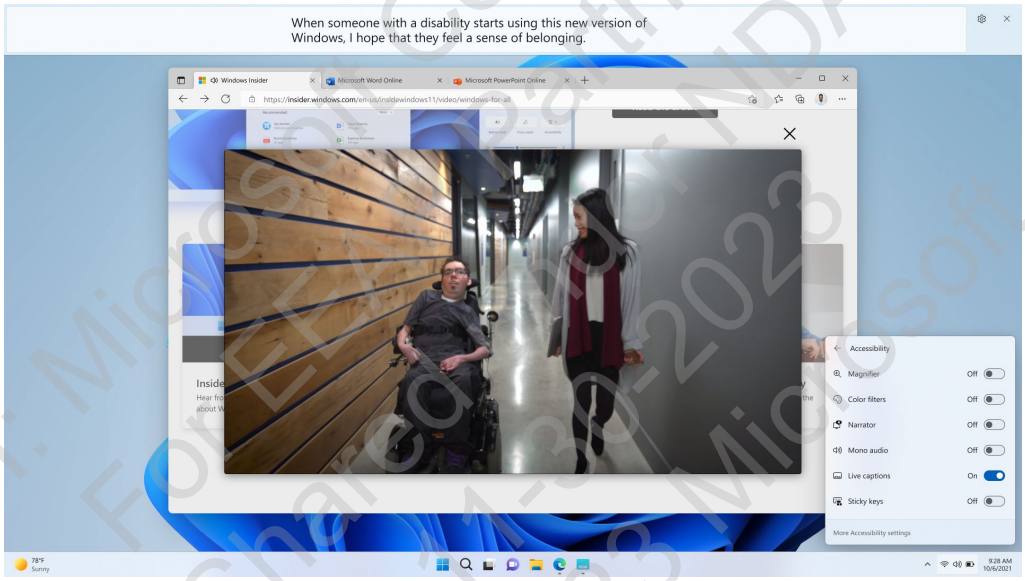
Enabling Natural and Intuitive Voice Experiences (Live Captions & Voice access)

Live captions and Voice access are new features being introduced in Windows 11 SV2. They will run on all Windows 11 SV2 PCs and process speech locally using CPU processing capability.

If there is a supported NPU in the system and Microsoft Effect Pack is present and opted-in correctly, Live Captions and Voice access will offload the speech models to run on the NPU for improved battery life, performance, and reduced CPU and system memory usage.

Live captions

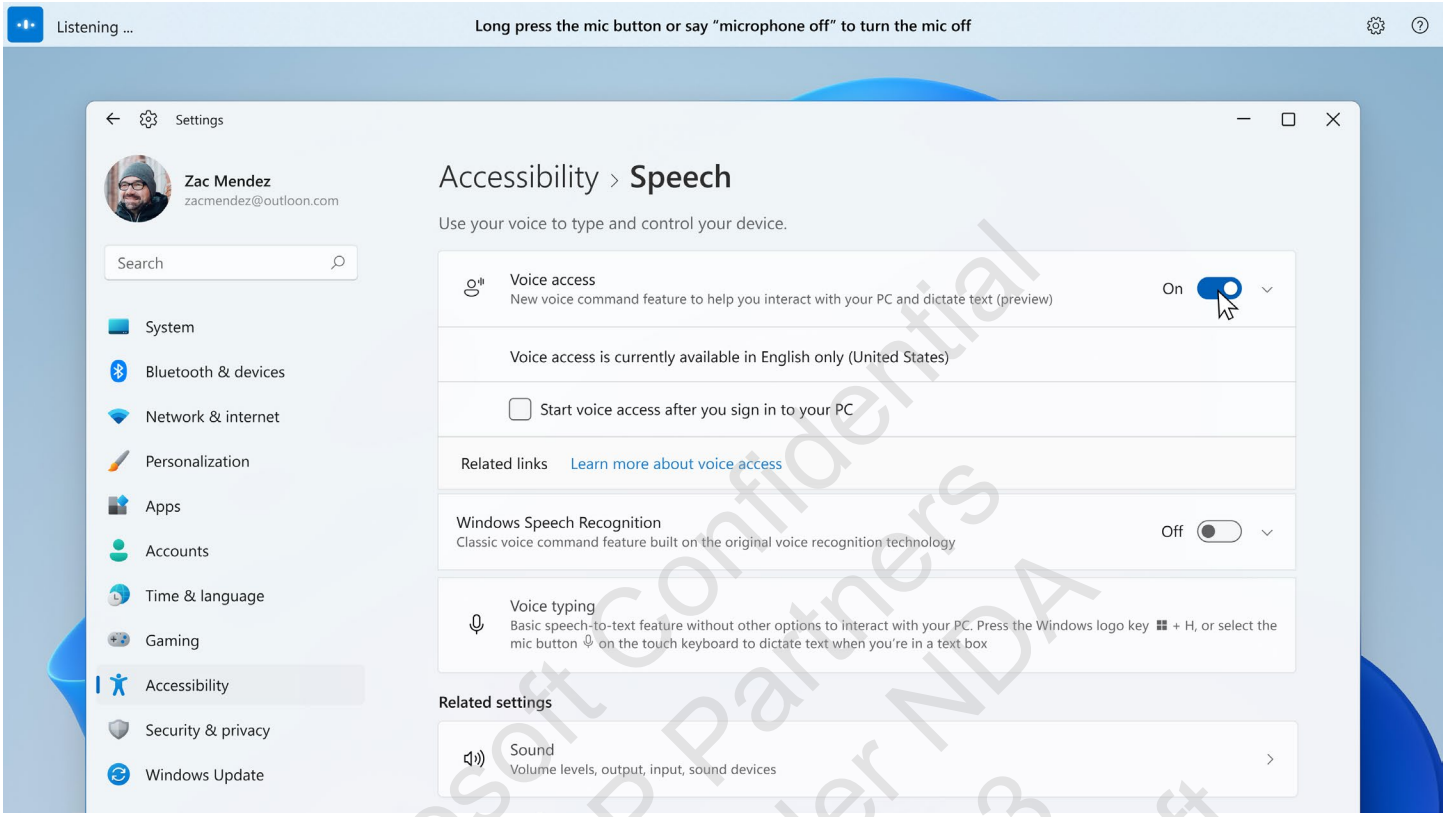
Live captions will help everyone, including people who are deaf or hard of hearing, better understand audio by viewing captions of spoken content. Captions are automatically generated on-device from any content with audio. Captions can be displayed at the top or bottom of the screen, or in a floating window. The caption window can be resized, and caption appearance can be personalized by applying or customizing a caption style. Microphone audio can be included, which can be helpful during in-person conversations. Live captions supports English (U.S.) content.



Live captions can be turned on with the WIN + Ctrl + L keyboard shortcut, or from the Accessibility flyout under Quick Settings. When turned on the first time, Live captions will prompt the user to download the required speech models to enable on-device captioning.

Voice access

Voice access is a new experience that enables everyone, including people with mobility disabilities, to control their PC and author text using their voice. For example, Voice access supports scenarios like opening and switching between apps, browsing the web, and reading and authoring mail. Voice access leverages modern, on-device speech recognition to accurately recognize speech and is supported without an internet connection. In Windows 11 SV2, Voice access supports English-U.S. language only, so the Windows display language should be set to English-U.S., otherwise voice access may not work as expected.

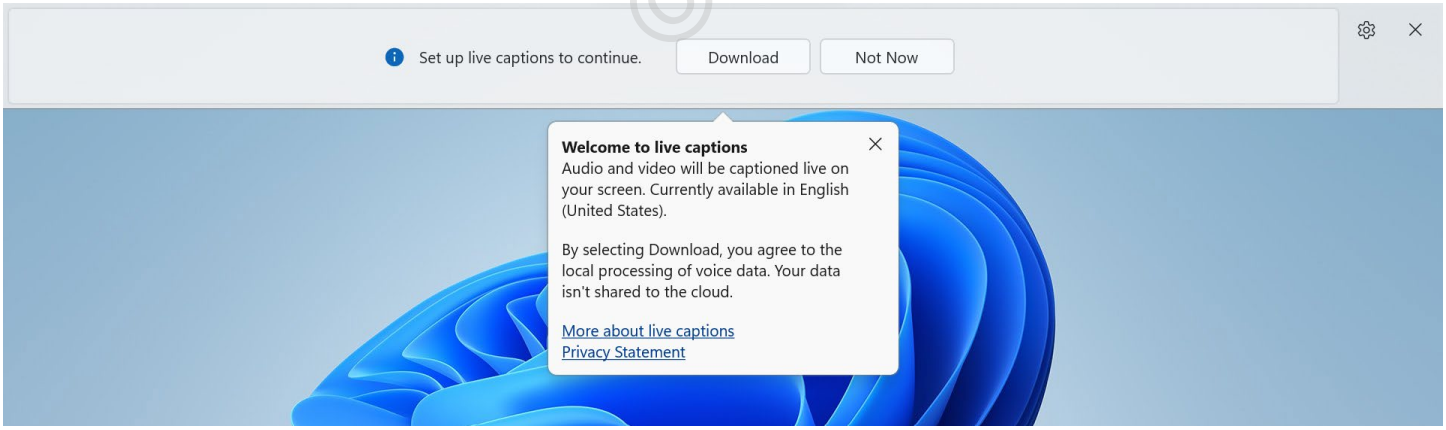


Additional information, including an introduction to common voice commands and the complete set of supported commands, is available in the voice access commands list.

Speech Recognition Language Models

Live Captions and Voice access use local processing capabilities of the device to use speech models that are deployed on the device upon first use. There are distinct speech models for CPU and NPU processing. Speech models are packaged in MSIX format and distributed and serviced to Windows systems using the Microsoft Store. The speech model for NPU is a larger and a more advanced model than the speech model used on CPU. Speech model packages are specific to a language and are approximately 100MB in size. Installed models are updated for quality improvements through the Microsoft Store

Upon the first run of either Live Captions or Voice access, the user will be presented with a dialog to download the speech model package as shown below.



Validating Natural and Intuitive Voice Experiences

The Azure Speech: Edge Devices specification provides OEM device makers with best practices for device design and a validation framework to build speech-enabled devices. We recommend that you review the specification and use the tool chain to optimize your devices for the best speech experience on Windows.

There are no additional requirements beyond the Azure Speech: Edge Devices spec to deliver a good experience for these features.

Additional Important Resources

Real Time Communication (RTC) Hardware Guidance - CY2022-2023 & CY2024

Windows 11 SV2 - Windows Studio Experience & Scenario Test Guide

Windows Studio Overview & OEM Extensibility for Cameras

Windows Studio Overview & ISV Integration Guidance

Frequently Asked Questions

Q: When will Microsoft enable developer access for 3rd party apps to leverage the NPU?
A: Refer to BUILD 2023 to learn more about ONNX Runtime and Olive toolchain investments: Unlocking the end-to-end Windows AI developer experience using ONNX runtime and Olive - Windows Developer Blog

Q: Will Voice experiences (Live Captions, Voice access) require internet connection?
A: No. Live Captions and Voice access will work offline.

Q: What is your timeline to support external cameras for the built-in Microsoft NPU effects?
A: This is not supported in Windows 11 SV2. We are considering adding support for external cameras in a future release.

Q: Can OEMs opt into specific effects, but opt-out of other effects on an NPU-system?
A: The opt-in model for camera effects is an 'all or nothing' model. The built-in camera AI effects cannot be opted-in individually.

Q: Will Windows Studio camera effects work on integrated front-facing USB cameras?
A: Yes, Windows Studio camera effects are supported on both integrated USB and MIPI-CSI cameras.

Q: Can I use both Custom INF and USB MSOS descriptors to opt-in the front-facing camera?
A: Opt-in of the front-facing camera can be implemented in both a Custom INF and USB MSOS descriptors at the same time, however the value written via the custom INF file will take precedence.

Q: Will the Customer Experience Review include assessments to validate the AI-powered experiences?
A: Yes, please reach out to your Microsoft account representative to obtain the latest CER methodology and documentation.