

OPTIMIZING TELECONFERENCE AUDIO SYSTEM DESIGNS



The Texas Department of Public Safety uses four digital signal processing boxes in its Emergency Operations Center to power 29 microphones, provide local sound reinforcement across six zones, enable audioconferencing, and provide audio for videoconferencing. This article provides tips to effectively design similar systems that require the use of multiple digital signal processing boxes.

Linking multiple audio DSP boxes to act as a single virtual system for large teleconference system designs looks simple, but can present serious implementation challenges. These five tips will make the most of your multi-unit system designs for video and audio teleconferencing applications.

by Kelly Hannig, CTS-D

Have you designed a large system for video or audio teleconferencing that looked straightforward and simple on paper, but turned out to be a configuration nightmare after it was wired up? Or perhaps you've even had to re-design the entire system to make it work properly? If you answered yes to either of these questions, help is on the way.

By using these basic connection and process flow ideas, you can simplify general routing requirements

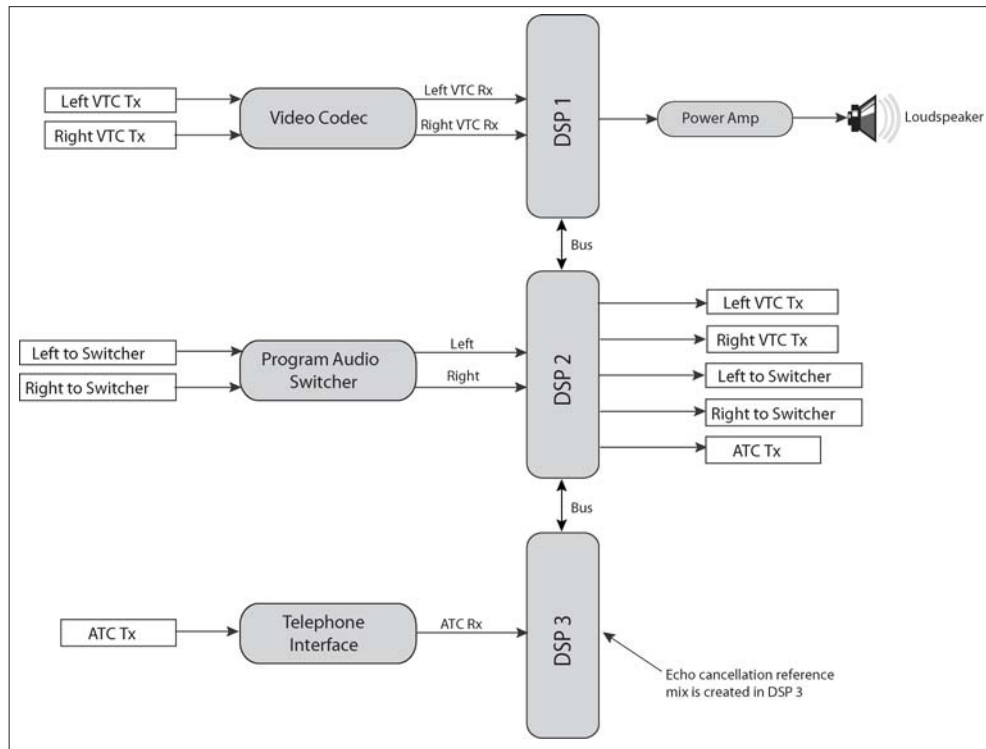
in the room, make it easier to manage echo cancellation references, and get the best system echo cancellation performance when using dynamic feedback controllers in a conferencing environment. The following five design tips will help you optimize multi-unit audio DSP systems for video and audio teleconferencing applications using the internal audio data bus more efficiently.

Tip #1: Simplify echo cancellation reference management by connecting all far-end audio sources and local program audio to the same physical DSP unit.

Echo cancellation, the enabling technology for teleconferencing, is simply the process of preventing unwanted audio from passing through a microphone channel to the far end. For example, we want audio from talkers in the local room to pass to the far end, but we don't want that same audio coming from the far end to return to the far end

Kelly Hannig is a field engineer for ClearOne Communications and provides design review support for consultants and system design engineers as well as support for on-site deployments of ClearOne conferencing products. He can be reached at Kelly.Hannig@ClearOne.com.

Fig. 1. A system layout that is not optimized for cancellation reference management.



To achieve echo cancellation, we must create a mix of the audio we don't want to pass to the far end.

because that's the "echo" we want to cancel. We also don't usually want local program audio to pass to the far end through an open microphone channel. Instead, it should be routed directly from the source to the far end.

To achieve echo cancellation, we must create a mix of the audio we don't want to pass to the far end. We call this the "reference mix" because echo cancellers on the microphone channels use it as a reference to create a signal with equal amplitude but opposite phased voltage. This signal is then applied at the right time to audio coming from the microphone to cancel out the audio that was in the reference mix (far end and local media). Audio not in the reference mix (local talkers' audio) is allowed to pass to the far end. (Important note: For best echo cancellation results, audio used in a reference mix should be post-process audio. In other words, the cancellation reference mix should be a sample of the signal being sent to the power amps.)

Fig. 1 illustrates a common audio flow design using multiple linked DSP units. Two specific inefficiencies exist in this design. First, it uses more audio paths than are necessary to create and distribute echo cancellation references. Second, it requires too many audio paths to create the mix-minus audio feeds for the far-end interfaces (telephone interface and video codec) and to route received far-end audio and transmitted local audio. These inefficiencies can lead to integration problems because audio paths are a finite resource. When used up, they require additional external wiring to make the system work properly.

Fig. 1 shows a system layout that is not optimized for cancellation reference management purposes because, again, it uses up unnecessary audio paths – a finite resource – to get all the needed cancellation reference sources together. In this design, audio from the telephone interface is connected to a different DSP box than the audio from the video codec, while program audio (non-speech audio sources that are also heard in the same room, such as a CD, VCR, TV, or DVD) is connected to a third DSP box.

If we choose to create the reference mix on the DSP box connected to the telephone interface, and if we choose to keep all audio separated, we tie up four busses just to move the codec L and R audio and the program L and R audio to the DSP with the telephone interface. Then we need to use another bus to pass the reference mix among the linked DSP units. We've now tied up five busses to create and distribute the reference mix. There's got to be a better way, right?

By doing a simple rearrangement of the physical connections, we can easily create a correct reference mix using only one bus instead of five.

Fig. 2 shows an audio flow design that is optimized for echo cancellation purposes. The far-end audio devices (telephone interface and video codec) and the local program audio are connected to the same physical DSP unit. This allows the reference mix to be created in that DSP unit. The reference mix is then placed on a single bus for distribution to all other linked DSP units.

Tip #2: Simplify management of audio feeds to the far end by connecting audio inputs and outputs of video codecs and telephone interfaces to the same DSP unit.

We can simplify feeds to far-end devices and external program switchers by arranging physical connections as illustrated in Fig. 3 on page 46. We don't have to jump onto any of the linked DSP units' busses to move local program audio to the far end or to create our mix-minus for the telephone interface/video codec cross-feeds. These can all be done on the DSP unit connected to these devices, without using linked audio busses.

Again, the advantages of connecting this way are fewer linked audio busses and simple management of the far-end device mix-minus cross-feeds.

Tip #3: Use the mix-minus feature of linked DSP audio busses to simplify mix-minus loudspeaker zone management.

Large audio and video teleconferencing spaces typically use local speech reinforcement to enable all participants to hear talkers in the same room.

Mix-minus feeds to loudspeaker zones are used to allow more gain before feedback in

Fig. 2 – An audio flow design optimized for echo cancellation.

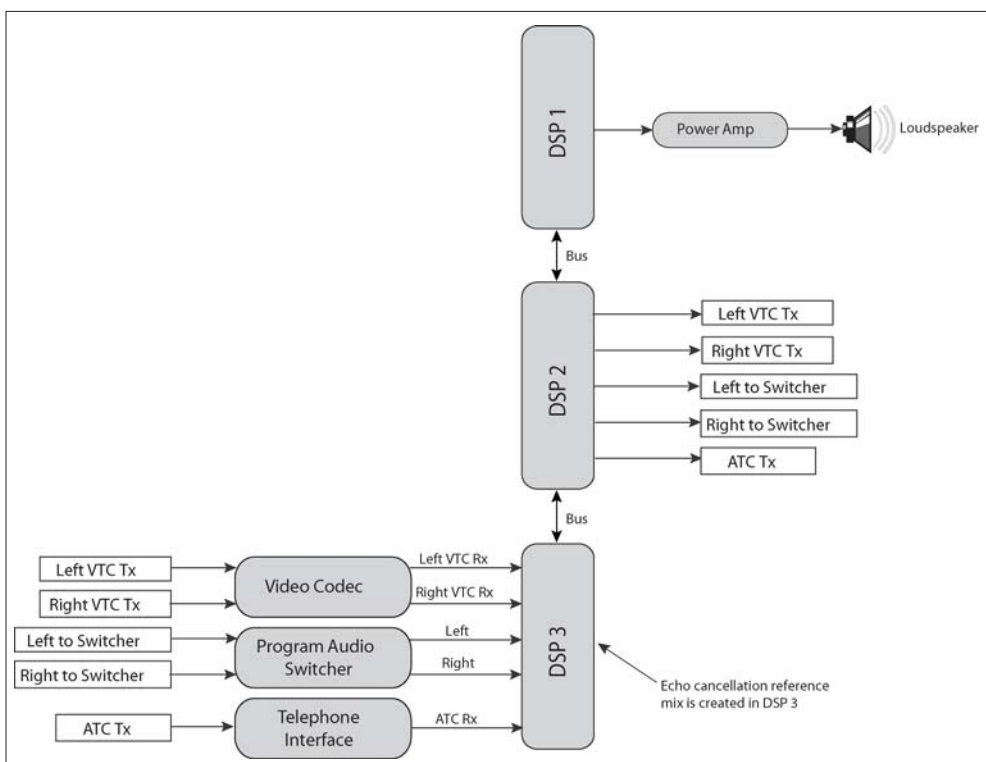
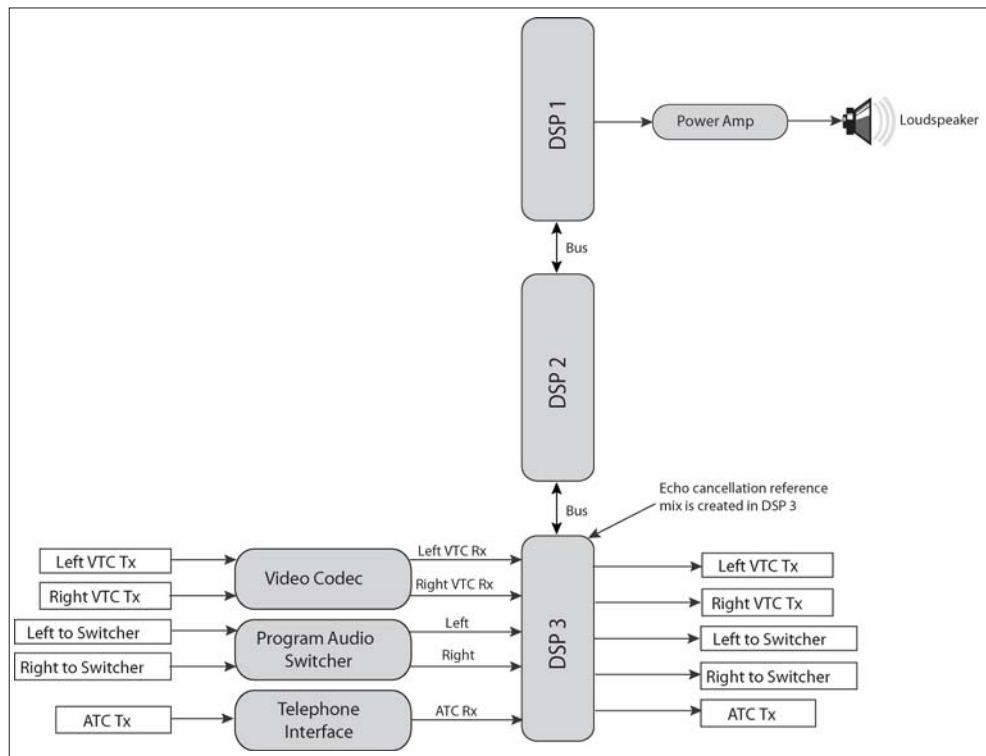


Fig. 3 – Simplified connection to far-end devices.



For best echo cancellation results, audio used in a reference mix should be post-process audio.

local speech reinforcement applications. For example, all microphones physically located under a specific overhead loudspeaker zone will be routed to all loudspeaker zones except the one directly above them. Therefore, local speech audio coming from the overhead loudspeakers of a specific zone contains a mix of all speech audio minus the audio from the microphones located in that loudspeaker zone.

Some DSP boxes have a mix-minus feature implemented on their linked audio busses. This is a very powerful feature, but users must understand that the mix-minus feature of linked DSP audio busses is not the same thing as the mix-minus feeds to loudspeaker zones. Mix-minus operation of linked audio busses is defined by the DSP unit's design, whereas mix-minus feeds to loudspeaker zones are created by the system designer using the DSP's matrix mixer/router. The following example illustrates how to use both of these mix-minus functions to simplify zoned speech reinforcement systems.

The layout in Fig. 4 is a basic mix-minus speech zone design, but it uses more linked audio busses than are necessary. In fact, it will require at least eight linked audio busses. To group microphones located in loudspeaker Zone E and send them to the first DSP unit will require the use of an audio bus. Similarly, mics located in loudspeaker Zone F will need to be grouped and placed on a separate audio bus. The same requirement exists for mics located in Zones G-L.

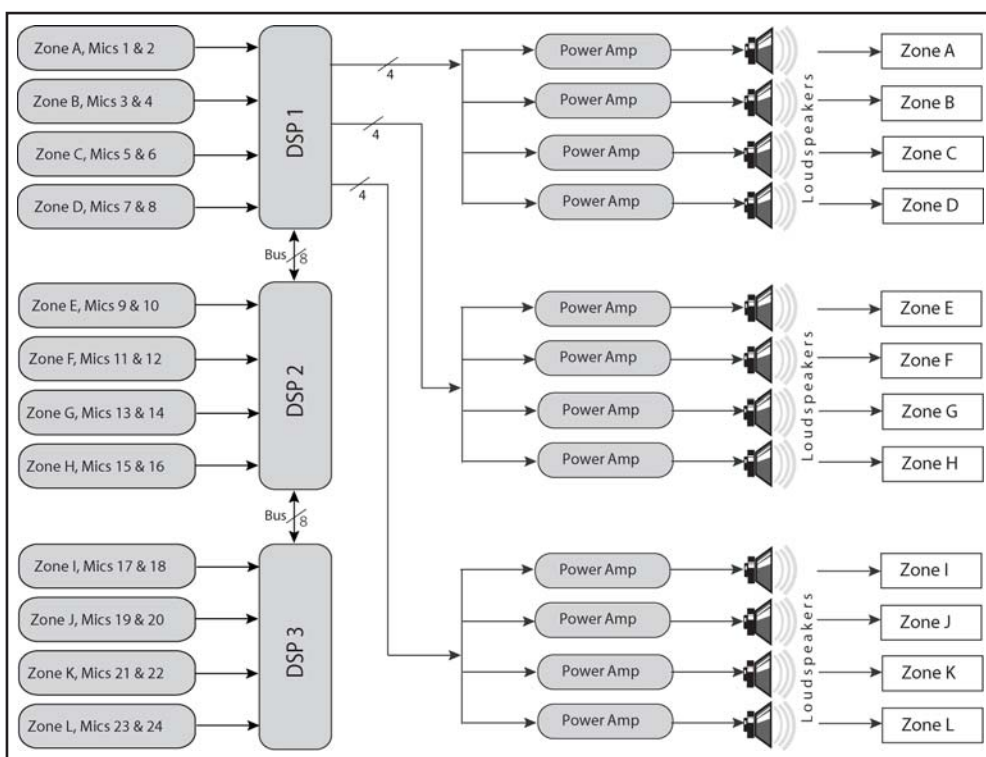


Fig. 4 – Basic mix-minus speech zone design.

In Fig. 4, mics located in loudspeaker Zones A-D don't need to be placed on external busses because they're connected to the DSP unit feeding all the loudspeaker zones. They can be routed directly to the required zones through the matrix of the DSP unit they're connected to. This is the key to using the mix-minus property of linked DSP audio busses.

By using DSP boxes that have a mix-minus feature on their linked audio busses, a simple re-grouping provides the same end result with only one linked audio bus instead of eight (see Fig. 5). While not every design falls as neatly into place as this one, the same concept can be used to simplify most designs.

Fig. 5 illustrates optimized connections for a mix-minus operation. Note that microphones located in a specific loudspeaker zone are connected to the DSP unit that feeds the same loudspeaker zone. For example, mics 9 and 10 located in loudspeaker Zone E are now connected to the DSP that feeds the power amp for Zone E loudspeakers.

It's important to understand the basics of a linked DSP audio bus mix-minus operation before applying it to a mix-minus zoned loudspeaker system.

Assume that all three DSP units illustrated in Fig. 5 place all their mics (1-24) onto the same linked audio bus. Mix-minus bus operation means that each DSP unit sees all mics placed on the single bus except the mics it placed on that bus. In other words, any given DSP unit looking at a specific bus will see a mix of all the audio on that bus minus the audio it placed on that bus. DSP Unit 1 will see mics 9-24 but not mics 1-8. DSP Unit 2 will see mics 1-8 and 17-24 but not mics 9-16. DSP Unit 3 will see mics 1-16 but not 17-24.

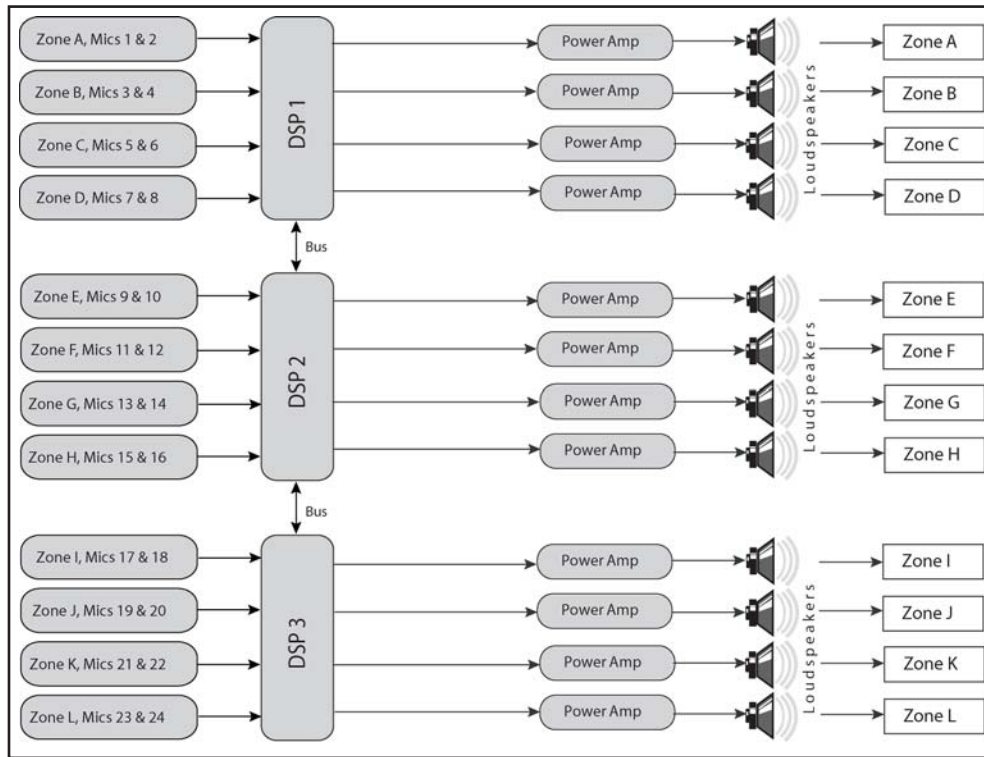
With that background, we can create a mix-minus loudspeaker zone feed. DSP 1 uses its matrix to feed mics 9-24 from the linked audio bus to loudspeaker Zones A-D. DSP 1 also uses its matrix to directly feed mics 3-8 to Zone A, mics 1, 2 and 5-8 to Zone B, mics 1-4, 7 and 8 to Zone C, and mics 1-6 to Zone D. Each loudspeaker zone fed from DSP 1 now contains mic audio from all zones except its own. The discussion for DSP Units 2 and 3 is identical.

Tip #4: For optimum echo cancellation results, avoid routing far-end audio through dynamic feedback controllers.

Dynamic feedback controllers are sometimes used when local speech reinforcement is required during an audio or video teleconference. Special consideration of audio paths is needed to achieve the best echo cancellation performance.

Fig. 6 on page 49 shows a common design using a dynamic feedback controller, which degrades echo cancellation performance in a conferencing situation. This is because echo cancellers compare audio returning from a room with the original reference mix. This comparison identifies what the room is doing to the referenced audio (far-end audio) in

Fig. 5 – Optimized connections for a mix-minus operation.



terms of acoustic absorption, delay, etc. The echo cancellers then make needed adjustments to adapt to changing room conditions. From the echo canceller's point of view, a dynamic feedback controller makes the room appear to be changing more than it really is, which degrades echo cancellation performance because the echo canceller is trying to adapt to a false acoustic "picture" of the room.

Fig. 7 on page 49 shows an optimized design for use of dynamic feedback controllers, which separates the signal paths for optimal echo cancellation performance. Audio from local microphones is routed to the dynamic feedback controller and then fed to the loudspeaker zone. This is normal for use with local reinforcement. The difference is that the audio received from the far end (and local program audio) doesn't pass through the dynamic feedback controller on its way to the loudspeakers. Therefore, the echo cancellers aren't presented with a false acoustic picture of the room's effect on far-end audio, which allows echo cancellers to converge more accurately and rapidly.

(Note: Fig. 7 illustrates use of dynamic feedback controllers with a single loudspeaker zone. When using multiple zones in a conferencing environment, each zone must have its own dynamic feedback controller.)

Tip #5: Improve echo cancellation results by using an optimized process flow between a microphone and the mixer.

Drag-and-drop architectures are appealing because of their flexibility. However, with increased flexibility comes increased opportunity to make mistakes in process flows, especially when using echo cancellers required for conferencing applications.

Fig. 8 illustrates an optimal echo cancellation process flow for a mic input channel in a conferencing

environment. The echo-canceller is strategically placed for best performance.

The first stage, gain, is straightforward, but notice that automatic gain control (AGC) functions aren't implemented here. Placing an AGC function prior to the echo canceller and noise canceller degrades their performance by giving a false view of the room.

The acoustic echo canceller (AEC) is next. It must be placed as close to the room as possible to accurately respond to real changes in room conditions.

Following the echo canceller is the noise canceller (NC). Like the AGC function, if this were placed prior to the echo canceller, the echo canceller would see a false acoustic picture of the room.

The mute function is applied after the echo cancellers. If it were applied before, the cancellers wouldn't be able to continually adapt to changes in the room. (Note: When using push-to-talk microphones, make certain their control mutes after the echo canceller and noise canceller. Otherwise, echo and noise may be heard for the first few seconds after un-muting while the cancellers are re-converging.) Filtering and AGC functions are also applied after the echo cancellers.

After processing a mic's audio for optimal echo cancellation results, the audio is finally presented to the DSP's mixer function. **AV**



Fig. 6 (left) – A common design using a dynamic feedback controller.

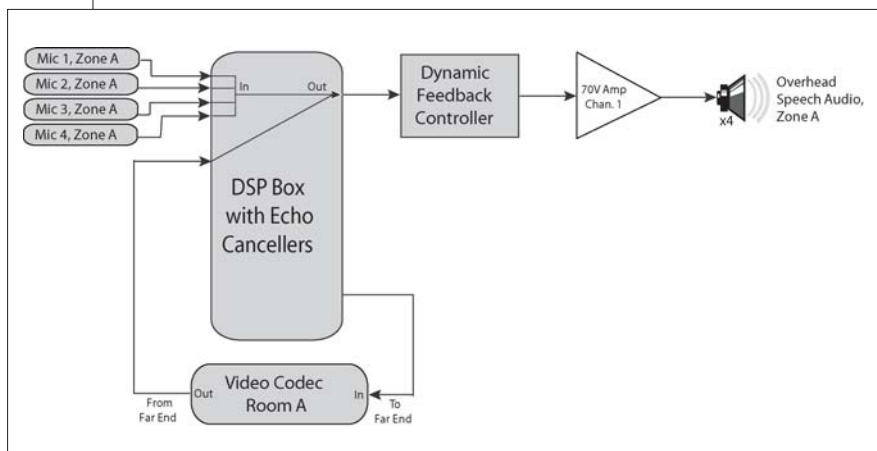


Fig. 7 (right) – An optimized design for using a dynamic feedback controller.

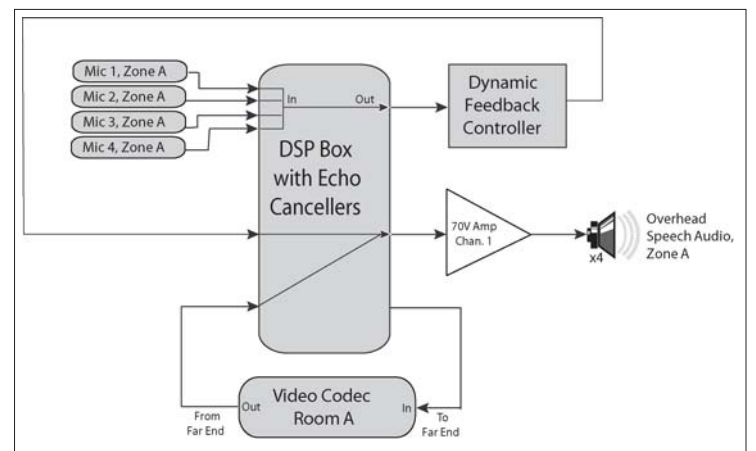


Fig. 8 – Optimal signal flow for the echo cancellation process.

