VISUAL COMMUNICATION SYSTEM-TECHNICAL DOCUMENTATION

PCS Series Superb Sound Technology

PCS-G60 All
PCS-XA55 All
PCS-XA80 All
PCS-XC1 All
PCS-XG55/XG55S All
PCS-XG77/XG77S All
PCS-XG80/XG80S All
PCS-XG100/XG100S All
PCS-XL55 All
Introduction

The Sony Visual Communication System (hereafter referred to as PCS *) realizes two-way communication using MPEG-4 AAC stereo sound with a high sampling frequency of 48 kHz as a standard feature. This covers a frequency band exceeding the 44.1 kHz sampling frequency of compact discs (CDs), which are known for their high-quality sound. The system is equipped with external input terminal(s), enabling music to be delivered in high-quality stereo sound. It is also equipped with a wideband stereo echo canceller. With these features which are incorporated into an extremely compact package, the PCS provides realistic, clear and natural sound appropriate for high-quality video.

This document describes the echo canceller technology and multi-rate signal processing technology which are indispensable for realizing these features.

*1: The supported models are referred to on the front cover. As for the PCS-XC1, see also endnotes on the last page.
*2: The echo canceller works at an internal sampling frequency of 32kHz.
Stereo Echo Canceller Technology

The PCS products (other than the PCS-XC1) are equipped with a wideband stereo echo canceller.

What Is The Acoustic Echo Canceller?
In videoconferencing, for example, in the case of two-way hands-free communication, the voice of a participant is sent to the remote site. The audio is output from the loudspeaker at the remote site, where it is picked up by the microphone and sent back to the loudspeaker at the local site. In other words, when you speak into the microphone, you hear your own voice from the loudspeaker as an echo. Known as acoustic echo, this phenomenon is considered annoying and can disrupt communication. In the worst cases, howling is created by feedback, making communication impossible. Therefore, acoustic echo canceller technology that eliminates sound from the loudspeaker and prevents it from being sent back to the remote site is indispensable for realizing full-duplex communication in two-way communication systems using loudspeakers and microphones.

Difference Between Monaural And Stereo
Adaptive filtering technology, which is at the core of the echo canceller, calculates the characteristics between loudspeakers and microphones and performs processing to cancel the sound of the loudspeaker picked up by the microphone.
For conventional monaural audio, systems consist of one loudspeaker and one microphone, which means only one adaptive filter is required because there is only one relationship. For stereo sound, systems consist of two (right and left) loudspeakers and two (right and left) microphones, meaning four adaptive filters are required because there are four relationships. Compared to the monaural echo canceller, the stereo echo canceller needs to perform four times the amount of calculation on a conceptual basis. Sony’s unique technology overcomes this difficulty.

PCS’s Stereo Echo Canceller
The PCS incorporates two sets of high-performance DSP with large-capacity memory. Building upon the high-end model PCS-HG90’s technology with a new algorithm developed, along with optimized processing of DSP, the PCS achieves a compact size and natural-sounding audio that compares favorably with the PCS-HG90, without affecting the performance of the stereo echo canceller, which requires an enormous amount of processing.
Multi-rate Signal Processing Technology

One way to reduce the calculation amount is to use multi-rate signal processing technology, which allows downsampling to lower sampling frequencies by splitting the frequency band for processing. For example, if the sampling frequency is halved by downsampling, the processing amount of the adaptive filter is halved.

It is known that the quality of the adaptive filter deteriorates if aliasing occurs during downsampling. An anti-aliasing filter is applied before downsampling to prevent this aliasing, but the characteristics of the filter affect the sound quality. When using the "maximally-decimated" method that can most reduce the calculation amount, one of the filter characteristics is that the stop band starts from the border with the adjacent split bands, thereby sacrificing part of the pass band to obtain sufficient attenuation within the stop band. As a result, when band splitting/synthesis is performed, notches are formed around the borders between adjacent bands, which adversely affects the sound quality.

To Realize Natural Sound

The PCS, which is designed with sound quality as the top priority, also addresses this problem. To solve the problem of the maximally-decimated method, an oversampling method is adopted. (See Figure 1.) Although the calculation amount with this method becomes about twice as large for adaptive filtering calculation, the stop band can be roughly doubled, which makes the boundary between adjacent bands flatter. In addition, aliasing is prevented. As a result, more natural, clearer sound with less distortion compared to the maximally-decimated method is realized.
Fig.1: Comparison with Maximally-Decimated Method

Band-by-band Characteristics

Maximally-Decimated method

Overall Characteristics

Oversampling method
New Technology For High-performance Echo Canceller

 Improved Response to Sound Field Changes

“Sound field changes” refers to the changes in characteristics that occur between microphones and loudspeakers when rearranging microphones and/or changing the amplifier volume. The PCS’s echo canceller uses an innovative residual echo suppression technology and a new reinforcement learning algorithm for the acoustic environment, which improve the performance for returned echo and realize quicker echo reduction. (See Figure 2.)

 Fig.2: Response To Sound Field Changes

 Enhanced Double-talk Suppression

“Double-talk” refers to the situation in which both parties speak at the same time. The PCS’s innovative residual echo suppression technology significantly reduces the residual echo, resulting in a more natural, comfortable conversation. (See Figure 3.)

 Fig.3: Echo Suppression During Double-talk Period
Improved Muffled Sound*3
In an environment where a speaker is speaking into a microphone from a distance, or speaker’s voice cannot be directly input to the microphone due to obstacles (e.g. laptop PC’s screen) which block out the voice, high frequency component is lost and it may cause a muffled sound which is difficult to hear. With the new echo canceller, the muffled sound has been much improved by a new technology which produces an enhanced audible sound. (See Figure 4.)

Fig.4: Improved Muffled Sound

Enhanced Audio Quality in Noisy Environment*3
When a microphone is placed near a projector fan or in a similar environment, noise is more likely to be input to the microphone and it may be difficult to hear what a participant at a remote site is saying.
New echo canceller that has a noise level estimation processor and suppresser can produce a clearer sound even in a noisy environment. (See Figure 5.)

Fig.5: Enhanced Audio Quality in Noisy Environment
Adaptive Voice Pickup Range*3

During videoconferencing, it may be hard to hear what a speaker speaking into a microphone from a distance is saying compared to vocalizing near the microphone.

New echo canceller allows you to set the pickup range and sound quality for the microphone according to your environment. (See Figure 6.)

Fig.6: Adaptive Voice Pickup Range

Solution to Long-latency Speakers

Flat-panel TVs such as LCDs and plasma displays perform a variety of internal video signal processing to reproduce a high-quality picture. However, this video signal processing can result in a delayed picture output (150 ms or more in some cases), causing the audio and video to become out of sync. To prevent this, their TVs insert the same amount of latency (lip sync) to the loudspeakers’ output for video and make them synchronize.

However, conventional echo cancellers cannot eliminate echo when long latency speakers are used. They attempt to recognize the latency as an elongated echo. But the echo time including the latency is too long to cover with the adaptive filtering, hence the echo cannot be fully eliminated.

The PCS’s echo canceller, however, resolved this problem. The new echo canceller has a latency amount following control which ensures the echo canceller to operate properly and to eliminate echo reliably, even when the PCS is used in conjunction with high-quality televisions with long-latency speakers. *3

*3: These features applicable to:
PCS-XC1: Ver.1.0 and later
PCS-XG77/XG77S: Ver.1.4 and later
PCS-XG100/XG100S: Ver.1.4 and later
Notes on PCS-XC1

- The PCS-XC1 supports monaural audio only, not stereo.
- A sample frequency is 48 kHz. The echo canceller works at an internal sampling frequency of 24 kHz.
- No external audio input terminal is equipped.

IPELA is a trademark of Sony Corporation.