



THE CURE FOR CLIPPING: SPEECH IN COMPLEX NETWORKS

AN APPLICATION NOTE

Jeff Rodman

Fellow/CTO

October 17, 2003

jrodman@polycom.com

© 2003 POLYCOM, INC. ALL RIGHTS RESERVED.

POLYCOM, THE POLYCOM LOGO, SOUNDSTATION AND VIEWSTATION ARE REGISTERED TRADEMARKS, AND SOUNDSTATION VTX 1000, VORTEX AND iPOWER ARE TRADEMARKS OF POLYCOM, INC. IN THE UNITED STATES AND OTHER COUNTRIES. ALL OTHER TRADEMARKS ARE THE PROPERTY OF THEIR RESPECTIVE COMPANIES.

INTRODUCTION

As telephone systems become more complex, problems with sound quality can be difficult to locate and fix. Clipping, the unwanted cutting off of parts of words or sentences, is especially hard to pin down because it can be subtle, it can come and go, and it can result from a variety of different causes in the system. This application note will help understand what clipping is, what causes it, and how to fix it.

CLIPPING: WHAT IT IS, WHAT CAUSES IT

“Clipping” in a speech network refers to anything that causes parts of words to be cut out of a section of speech (there is a second, very different effect sometimes called “clipping” or “flat-topping”, which is a form of analog distortion; that is not discussed in this paper). Clipping can occur for a wide variety of reasons, including:

- Network problems
- Half-duplex speakerphone
- High background noise level
- Amplitude switching or VAD (Voice Activity Detection)
- Calling through a multipoint bridge
- Mixed telephone/videoconferencing calls
- Network latency
- Dropped packets
- Jitter buffer over/underflow
- Far end using inferior equipment

As you can see, many of these can occur at more than one place in a network, especially in more complex networks. Because a channel carries voice all the way from the talker’s mouth to the listener’s ear, anything in between these two can be responsible for clipping: the talkers themselves, the telephone or speakerphone, the network or parts of the network, the listener’s phone, or the listener or their environment.

In spite of the number of potential causes, it is usually straightforward to isolate the cause of clipping, and to implement a solution.

Analog Networks

First we shall look at the simplest case, a plain analog phone call. Figure 1 shows this connection as occurring between two analog desktop phones over the public network.

THE CURE FOR CLIPPING

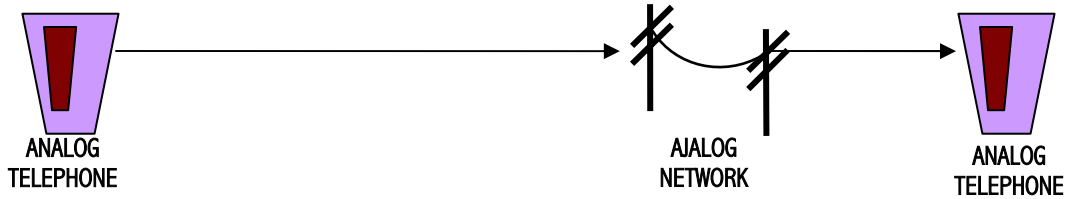


Figure 1. Simple analog connection

This configuration brings together two analog desktop phones, which are normally full-duplex instruments, with the highly reliable public switched network. Clipping rarely occurs in this case, because there are relatively few specialized functions occurring between talker and listener. Telephones connect the microphone and earpiece directly to the phone line, with no switching in between, so there's nothing to cause clipping here. The telephone network operates as a continuous channel between the two ends, so nothing can cause clipping here either. Right?

Not always. As telephone networks span greater distances, they employ an increasing amount of technology to perform their jobs efficiently. Most "analog" networks today are in fact digital networks with analog interfaces at the ends. These networks sometimes take advantage of the fact that much conversation is actually silence: when a person is listening, they are usually not talking, and this is valuable channel capacity that can be used elsewhere. A synthetic silence, often a low-level artificial noise signal, is inserted in its place, and the real silence is replaced with speech from another channel for a few moments (Figure 2).

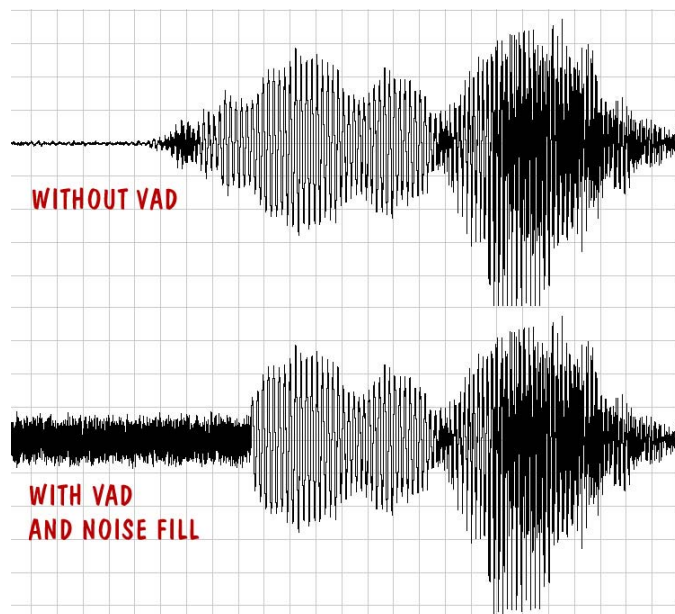


Figure 2. VAD and noise fill

By doing this kind of millisecond-by-millisecond processing on hundreds or thousands of conversations at once, a network can borrow enough silences to carry almost twice as many conversations on the same cable, and so reduce the cost of providing the service. We will return to this concept in discussing VoIP networks below.

Speakerphones and Clipping

Speakerphones are sometimes identified as the culprit when they are in fact innocent, or are being challenged beyond their capabilities.

In order for a speakerphone to not clip the conversation, it must be a true full-duplex system, meaning that audio must flow continuously from each end to the other. Conventional speakerphones do not do this, they operate by switching audio back and forth, and so cause clipping. Only speakerphones certified by the manufacturer as being true IEEE 1329 Class 1 full-duplex systems will give good non-clipping performance. Moreover, the instruments at both ends, not just one end, must be full-duplex systems, whether handset telephones or speakerphones. Note that the “speakerphone” function built into most handset phones does not meet this performance standard; even most systems labeled as “full duplex” are in fact half-duplex speakerphones most of the time unless the manufacturer clearly specifies their compliance with IEEE 1329 Class 1 specifications (sometimes not even then, but that’s a matter for another paper).

However, even the best full-duplex systems operate in a half-duplex mode for a few seconds at the very beginning of a call. This is because full-duplex speakerphones require a few seconds of audio in each direction to “train”, which is the process of analyzing the echo environment in the room and on the line, and constructing a computer model to accurately subtract echoes. Consequently, clipping in the first few seconds, although rarely noticed, is normal, and is not an indication of instrument or network problems.

The endpoint environment can also be a source of clipping; background noise in particular should be kept to a minimum. High levels of noise can confuse both full-duplex and half-duplex speakerphones, and cause them to switch the voice in and out unpredictably. Fan noise, air conditioner noise, overhead projectors, outside construction or traffic, or background conversations can contribute to this problem. If this is suspected, try moving to a quieter environment and see whether the problem goes away.

Another endpoint environment that can cause problems is one that has peculiar reverberation properties. Rooms with many reflective surfaces, such as large windows and whiteboards, can have unusually high reverberation. This will make a person who is three feet from a speakerphone sound like they’re fifteen feet away, and can also cause a full-duplex speakerphone to make occasional errors in its efforts to avoid clipping. So this situation is best avoided for a number of reasons when possible (see also Polycom whitepaper, “The BRAIN Model of Intelligibility in Business Telephony”).

VAD and VoIP Networks

In Figure 3, we see an IP system in which two IP endpoints are connected across the public network. This is done by converting the IP signals to standard telephone signals in a gateway, and is one of the most common configurations today as it allows companies the management and installation advantages of IP telephones while maintaining the reliability and accessibility of the public network. In this configuration, there are at least six places where VAD may be implemented and turned on, so these would be among the first to check if clipping is an issue.

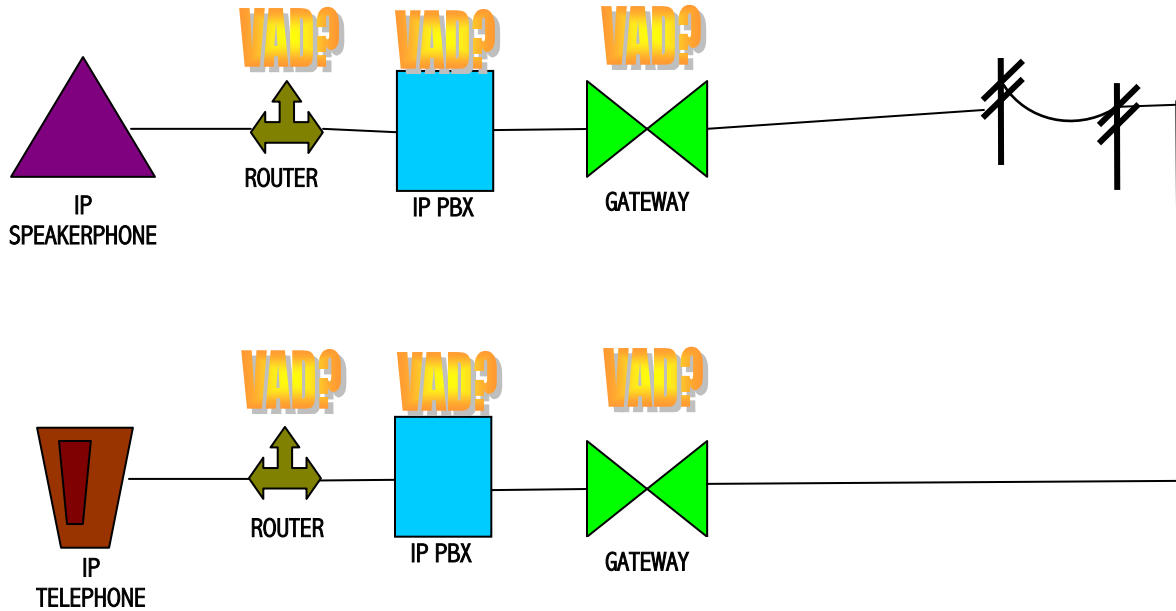


Figure 3. Simple VoIP connection

VAD is very common in VoIP networks because of the degree to which it can reduce network traffic. Some are shipped with VAD enabled by default. Because of other characteristics of VoIP networks, however, including the need to packetize the audio, VAD is often responsible for network-induced clipping. Consequently, disabling VAD should be one of the first steps in diagnosing a clipping problem. When disabling VAD, be sure to check every piece of equipment that the voice signal passes through. VAD is commonly found in switches, routers, gateways, call managers, and other network data systems, so getting a VAD-free path may involve more than one change in network configuration.

In the Cisco Call Manager, there are two instances of VAD; they must both be turned off. They are located as follows:

- Login
- Click Service
- Click Service Parameters
- Click the IP address of the Cisco Call Manager in question

THE CURE FOR CLIPPING

Click Call Manager from the left hand column
Scroll down to the "S"s. On CallManager 3.2 the parameters are in alphabetical order:

SilenceSuppressionSystemWide
SilenceSuppressionWithGateways

Also note that either a reboot (preferred) of the server or a restart of the services (both Call Manager and TFTP server) is required before the parameter change will take effect.

VAD problems can disguise themselves as speakerphone problems, as when clipping is heard with a speakerphone but not when a telephone is substituted. While this may be an indication of a speakerphone issue, it also can be caused by the fact that speakerphone audio is usually less loud than telephone audio, and so is not detected as quickly by the VAD function. VAD systems are often not optimized for use with speakerphones, so will sometimes perform less reliably. This is shown in Figure 2, in which the first part of the word has actually been cut off and replaced by synthetic noise; it has taken several milliseconds for the system to recognize that speech has started, but by then the beginning of the word is already lost.

Low speech level problems are often fixed by using a wireless microphone for quiet talkers, by moving the microphones closer to the talkers, or by adding more microphones in the room with a system such as the Polycom Vortex.

Packet Dropouts

VoIP networks have an added complication, which is that they do not recover from lost data packets in the same way as IP data networks. Data networks using protocols such as TCP/IP have built-in safeguards against lost or scrambled data, and achieve their high reliability by automatically requesting retransmission of a packet if it has not been received. This process can take hundreds or thousands of milliseconds, which is much too long for a voice signal. Consequently, VoIP networks use a different technique.

In VoIP networks, the first line of defense against lost data is network reliability. VoIP IP networks are usually architected with considerable care to minimize latency and data loss. The second line of defense, when a packet is lost, is "error concealment", any of a variety of techniques that minimize the perceptibility of the lost data. Even the best error concealment algorithms, however, can only conceal so much, and once three or four consecutive packets are lost, an audible dropout is unavoidable. Clearly, this has nothing to do with acoustical causes, full-duplex, or even VAD, but is the result of random data errors on the network. The end result, though, is a kind of clipping in which wanted parts of speech are clipped out, so it must be considered in any analysis.

PBX and Multipoint Bridges

A PBX or multipoint bridge can introduce clipping when it is used to bridge together more than two participants on a phone call. This is because bridging systems, including PBX's

with built-in bridges, operate by selecting the loudest talkers and switching off the rest. This function is a kind of VAD. PBX's commonly do this for three, four, sometimes six users, but beyond that specialized equipment is used. Such bridges are at the heart of conference systems in many companies and commercial conferencing services. This switching in and out can cut off the beginning and ending parts of words, introducing clipping. As with other VAD systems, speakerphones are more easily clipped by VAD than are handset phones, because of the lower signal levels they produce.

When diagnosing a clipping problem, be sure that the connection is point-to-point, and is not being placed through a conference bridge.

Latency

A long end-to-end time delay can cause the illusion of clipping because of the unnatural interaction that occurs. This is heard in some satellite communication systems, video systems, multipoint bridging systems, and some VoIP systems where the data delay is excessive. The time for a response can be long enough that the questioner decides to ask the question a second time, thinking that it was not heard, and this second question is started just when the answer starts to arrive from the far end. Both parties then talk and back off in turn, and several seconds of awkwardness ensue. This can sound like clipping because it is also very artificial. It is not clipping, however, because the audio is not interrupted, it just takes a long time to arrive.

Jitter

In digital IP networks, the time for a data packet to get from one end to the other is not tightly controlled, as it is in the public switched network. This delay can vary substantially even from one packet to the next, so a VoIP system requires a function called a "jitter buffer" to smooth out this data flow. When jitter buffers are too small, they can be overloaded by a packet with greater than normal timing variation; when this happens there will be a packet error, sounding just like a packet dropout (see above). The design of a jitter buffer is a delicate tradeoff because jitter buffers add latency. Too small, and more packet dropouts occur, while too large, and the latency is objectionable. The best solution to this issue is to be sure that the network itself has been designed to minimize jitter and latency, thus allowing a small jitter buffer.

Conclusion

Here is a short list of things to check when looking for the source of clipping, and a suggested step-by-step procedure to get at the root of the problem.

- Does the problem occur point-to-point?
 - *Multipoint bridges can cause clipping*

- Are both ends using handsets or true full-duplex (IEEE 1329 Class 1) systems?
 - *Inferior phones will cause clipping*
- Is VAD disabled through all parts of the network used by the voice connection?
 - *VAD often causes clipping*
- Does it occur more often with quiet talkers?
 - *Quiet talkers can challenge the best of systems*
- Is the background quiet?
 - *High background noise can cause systems to clip*
- Is there low latency in the connection?
 - *Long delays can sound like clipping*
- Is clipping occurring in the middle of words as well as the start and end?
 - *When not specific to the beginning and end of words, indicates data dropouts, not clipping*
- Has the speakerphone been given a few seconds to train?
 - *All speakerphones will clip until they have trained*

TROUBLESHOOTING

A good way to begin troubleshooting is to start with a simplified subset of the system. Add system elements back in one by one. Here is a good sequence to follow.

1. If the clipping problem is observed using an analog speakerphone, place an analog call from the speakerphone to an analog handset telephone over an analog connection. If the clipping problem persists, check the following:
 - a. Background noise level
 - b. Is the speakerphone a true full-duplex system, certified to IEEE 1329 Class 1?
 - c. Has the speakerphone been given several seconds of one-way speech in each direction to train?
 - d. Is the handset telephone operating in its "speakerphone" mode?
 - e. Is the call point-to-point, or is there a multipoint bridge?
 - f. Does clipping go away when talking more loudly? Try moving closer to the microphone, or using a wireless microphone.
2. If the clipping problem does not exist in the previous test, add parts of the system back in one by one, or monitor the signal, heading outward from the source module by module.

THE CURE FOR CLIPPING

Figure 4 is an example in which the investigation was started from the purple analog speakerphone. This was tied to a digital PBX, then connected to a VoIP network via a gateway. The green "NO CLIPPING" and red "CLIPPING" legends show where in the network clipping was noticed. In this case, the problem was found to be one missed VAD setting in router 1. When that was turned off, the clipping stopped.

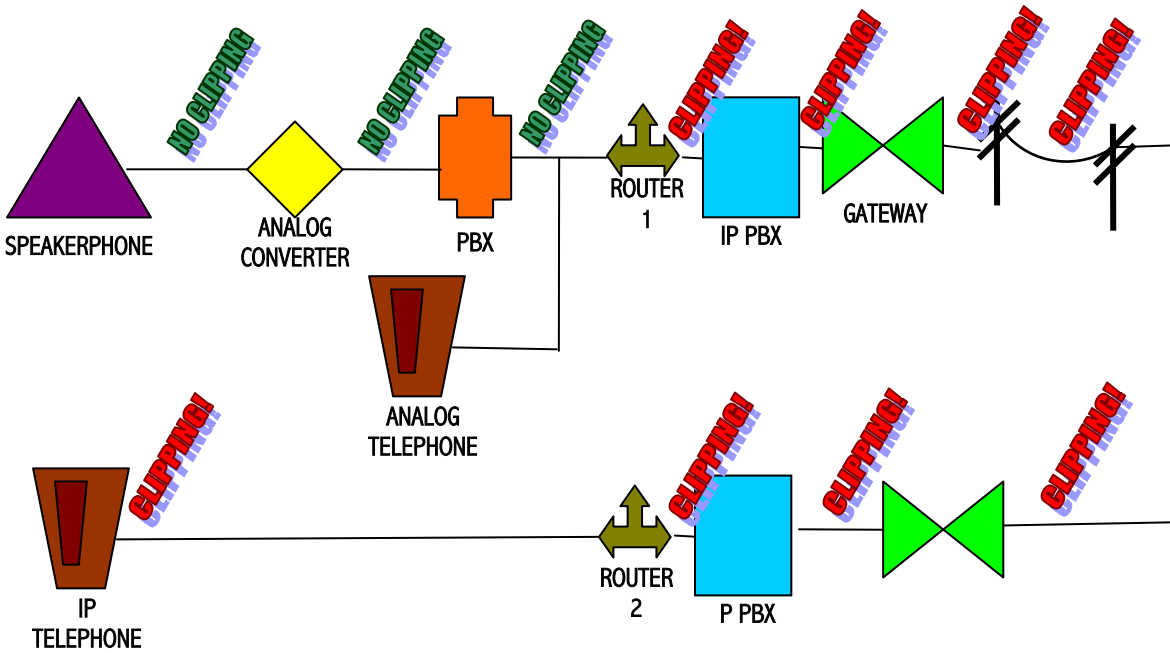


Figure 4.. Step-by-step examination of complex network