MVS

Jason Rohrer

October 27, 2005

Abstract

The goal is to develop a new and revolutionary product that combines video, mobile devices, peer-to-peer, and social networking. I thought about what has already been done in these areas and about what people might want to do that they currently cannot do. I came up with an product that, if designed and marketed properly, would certainly be revolutionary—in fact, it could completely re-weave the cultural fabric of our society. This idea bears very little resemblance to the "Veeker Plan" as discussed in the conference call on September 22, 2005.

I call this product the *micro-video station*, or MVS. This document describes the MVS in detail.

1 What already exists

1.1 Authoring video

People can author their own video using a wide variety of consumer devices. Some of these devices, like digital still cameras (DSCs, including those in camera phones), can produce relatively low-quality MPEG clips that are ready to be posted online or sent via email. Other devices, like digital video cameras (DVCs) produce high-quality, uncompressed videos in DV format that must be compressed and processed extensively before being posted online.

1.2 Editing video

A variety of consumer software exists for editing digital video. Some software packages, like the popular and affordable iMovie or the free Windows Movie Maker, focus on interacting with DVCs and editing raw DV data.

1.3 Sharing video

Services like YouTube allow people to post their personal videos online, without cost, for their friends or the world to see.

1.4 Consuming video

Video consumers can also use services like YouTube to discover and watch video content. Though YouTube supports video streaming (click-and-play, with no download wait), server bandwidth is apparently a problem, since videos tend to pause and stutter, especially during times of peak server load.

Other sources for videos, both legal and contraband, are P2P file sharing services. Most of these videos can only be viewed after a lengthy download wait.

2 What is missing

2.1 Pits in the landscape

The current video-sharing landscape is rather rough.

First, we have relatively-expensive DVCs (and their requisite FireWire cards), which are shooting high-quality video that is overkill for online sharing. As mentioned above, posting DVs online requires significant conversion and compression. Editing DVs in an uncompressed state is very data-intensive and slow, even on the fastest computers (let alone on the average consumer computer). Again, wasting CPU cycles editing uncompressed DV data, if the end-target is online sharing, is overkill.

Second, we have relatively-inexpensive and ubiquitous DSCs which shoot MPEGs that are perfect for online sharing. However, there are many obstacles to editing the resulting MPEGs before posting them. Most DV editing packages have trouble dealing with the lower-quality MPEG clips that can be authored with DSCs.

For example, in the version of iMovie that shipped with Mac OS X 10.2, MPEG clips cannot be imported at all. To edit a 15 second (1.3 MiB) MPEG movie from my DSC, I had to use a third-party utility to convert the MPEG into a 51 MiB DV file—the conversion operation took over 10 minutes. iMovie was then able to import and edit the DV file. After the edits, I was able to export a compressed MOV file from iMovie that was ready to go online. Figuring all of this out wasted several hours of my time and is probably out of reach for average consumers who want to edit MPEGs from their DSC. Furthermore, the two format conversions (MPEG to raw-DV and then DV to compressed MOV) resulted in substantial visual quality loss.

Some network-enabled DSCs, like camera phones, allow users to post videos online immediately after they are captured. One example is the EasyShare-One, the new WiFi-enabled DSC from Kodak, which shoots MPEG4s. The problem with the direct-upload approach is that videos must be posted in their rough, unedited state (perhaps after a bit of trimming, if the camera supports it) . This severely limits user creativity—not only is multi-clip assembly out of reach, but so is something as simple as adding sound track music.

In the video-consumption corner of the landscape, we still don't have click-and-watch video functionality, even from the services like YouTube that promise it. Even further out of reach is anything remotely like the instantaneous "channel flipping" that is possible with cable or satellite TV. Thus, "Internet video," despite all of the hype, still languishes far behind television in terms of the content-consumer's experience.

Despite all the talk about empowering consumers to create and distribute their own content (the

Internet, after all, is the great equalizer for mass media), we have essentially stuck the consumer with the ability to "broadcast" content with an experience-quality that is much lower than what is available to big media companies through cable TV. "Yes, you can finally broadcast your own show for free, but the viewing experience for your audience is going to be clunky and annoying."

2.2   What we *still* cannot do

Average consumers still do not have all-in-one technology which will enable them to:

1. shoot ready-to-share video clips;

2. edit and assemble multiple clips into a movie;

3. mix in a stereo sound track;

4. "broadcast" the finished product to audiences of unlimited potential size;

5. pass videos instantly to friends in face-to-face settings; and

6. watch the videos created by others with an instant, channel-flipping experience that rivals TV.

When we talk about empowering users to participate in mass media, we cannot simply think about how to give them access to TV broadcasting through cable access centers or about how to *mimic* TV using iMovie and the Internet. On one hand, a single cable station in each town cannot possibly fill the needs of a whole townful of potential video producers, nor can it enable a producer in New York to deliver videos to a consumer in California. On the other hand, the Internet was simply not designed for many-to-many video transmission.

Instead, we need to think about inventing an entirely new video medium—a consumer-centric video medium that is designed from the ground-up to allow each and every user to author videos that will potentially be seen by millions.

2.3   Technological longevity

If we consider the medium of television, we see a technology with extraordinary longevity, at least by modern standards. TV has been available to consumers for over 60 years, and during that time, the technology has changed very little. In fact, modern TV programs can still be viewed on some of the very first TVs that were brought to market, and modern equipment (such as digital cable receivers) can be easily connected to older equipment (like black and white TVs). The point is that the technology, the signal, and the entire medium were standardized when they were introduced and have not changed in any fundamental way since then. This technological stability is a solid foundation for a cultural revolution, and few would disagree that TV has shaped American culture during this century more than any other technology.

Comparing TV to a medium like the Web, we see a sharp contrast. Most "modern" websites will not render properly in older web browsers. Some of these websites would not readable *at all* in the browsers available for older machines. Instead of technological stability, we have an alphabet

soup of constantly-evolving standards and a frenzied, landfill-filling, 18-month hardware upgrade cycle.

If we want to shape culture in a profound and lasting way, we need a medium that will have the longevity of TV. Imagine trying to develop a Web-based technology that will still work and be useful in 60 years. Such a proposition sounds ridiculous, because *no one* is thinking long-term about the Web. Instead, Web innovations can be described as *hot today, gone tomorrow*. Thus, today's Web trends are at best insignificant blips on the long-term cultural radar.

The same holds true for any technology that is on a repeat-upgrade cycle. Consider video game consoles, like the PS2 or XBox. They will both be obsolete next year (with the release of the PS3 and XBox360), and the games available for them will be completely unplayable for most people within 10 years. Each new generation of game consoles has a significant impact on a new generation of game players, but none of them really leave a lasting, long-term term mark on society. These technologies become obsolete before they ever get a chance to seat themselves deeply into culture.

Thus, to create a truly new medium, we need a new kind of device with a technological stability inspired by TV. The *micro-video station*, or MVS, is a single device that will empower users to do all of the things they still cannot do from the list above. The most surprising thing about the MVS is that it will enable each and every person to reach a potential audience of millions *without* using the Internet at all. The second-most surprising thing about the MVS is that it is built on top of ten-year-old technology.

The rest of this document is devoted to describing the MVS.

3   The micro-video station

3.1   End-user's perception

An MVS is a hand-held device which will enable the user to create, edit, and share personal videos. First, I will describe what a single MVS can do in isolation. It can

- shoot digital video clips with stereo sound;

- assemble a series of clips into a movie;

- trim clips (frame-accurate cuts);

- browse through finished movies; and

- play finished movies (including fast forward-backward search).

Next, I will describe how an MVS can interact with other household devices. It can

- play movies on a TV screen;

- record incoming video signals, from TVs or VCRs, as clips; and

- transfer clips, movies, and music files to and from a computer.

Next, I will describe how an MVS can interact with another MVS in a face-to-face, friend-to-friend setting. It can

- mark certain movies as public and others as private;

- connect to another MVS using the supplied cable;

- browse and preview public movies on the other MVS; and

- transfer particular movies from the other MVS.

Two MVSs, when connected, can browse each other's public movie collections and transfer clips simultaneously. Thus, Alice can transfer a clip from Bob's MVS at the same time as Bob is browsing through or transferring clips from Alice's MVS.

Finally, I will describe how an MVS can interact with nearby MVSs in a stranger-to-stranger setting (for example, on a bus or in a shopping mall's food court). It can

- broadcast a particular movie or sequence of movies on a channel in the immediate vicinity;

- receive and display broadcasts coming from other MVSs in the vicinity;

- flip from channel to channel instantly; and

- save incoming broadcasts as clips with no loss of quality.

Figure 1 on page 10 shows various mock-up views of the MVS.

Note that, in the above description, the Internet is not mentioned at all. Instead of transferring movies through the Internet, we are passing them directly from friend-to-friend. The radio capabilities of the MVS also allow movies to pass from stranger-to-stranger, though strangers must occupy the same physical space. Thus, you can only directly share movies with family, friends, and strangers who have something in common with you (for example, strangers that frequent the same coffee shop as you).

Movies spread through the population via a true social network. Content is filtered by this network, and only the most interesting movies can travel far. On the other hand, a clip with mass appeal could potentially reach every MVS owner on the planet in a few days time.

"Ohmygod! You just *have* to see this movie I got from Sarah yesterday."

3.2 Technical details

This section describes the technology that makes the above end-user experience possible.

3.2.1 Shooting and editing video

Considering the importance of technological longevity, we need to chose a video format which offers sufficient quality (for display on a hand-held screen or a TV) but is already an open, widely-adopted standard. Though choosing particular format parameters will depend on a number of engineering constraints, something like MPEG1, 320x240, 30 FPS, and 1.5 Mb/s would be sufficient. By picking a particular video format and making a standard out of it, we can build a strong

and stable foundation for the content created and shared through the MVS (the example format given above, for instance, is viewable on almost every computer system manufactured in the last ten years). Also, choosing a single format will help to streamline the hardware that drives the MVS.

Assuming the example format given above, we can sketch out more technical details. The MVS is equipped with a zoomable lens and a 320x240 pixel CCD for capturing digital video. A hardware MPEG1 encoding chip converts the incoming, raw digital video into the MVS clip format.

The MVS is equipped with an internal flash drive for storage. A 2 GiB drive (similar to what is inside the new Apple Nano, which retails for $200) will store up to 3 hours of video at 1.5 Mb/s.

The screen on an MVS is a 32-bit color LCD with 320x240 pixels. The screen is used both as a viewfinder while shooting and a display while viewing.

Video capture begins whenever the red recording button is pressed, and ends whenever the button is pressed a second time—each captured segment is stored to the flash drive as a clip. Clips can be browsed using a thumbnail view. Selecting a particular thumbnail and turning the jog wheel flips through the key frames of the associated clip.

Whenever a video source is plugged into the A/V In port on the MVS, the incoming video signal overrides the input from the lens. Clips can be captured from the incoming signal just as they are captured from the lens.

Editing movies is accomplished by choosing an ordered sequence of clips, trimming each clip using the jog wheel, and saving the finished arrangement as a new clip.

3.2.2 Dealing with sound

Sound is captured in stereo along with each clip via the stereo microphone pair on the front of the MVS—the sound accompanying a given clip is referred to as the *ambient sound* throughout this document.

When clips are assembled into a movie, the loudness of their ambient sounds can be adjusted. Any MP3 music files that have been transferred to the MVS from a computer (or from another MVS) can be sequenced and trimmed to serve as the soundtrack for a given movie. Ambient sound from a given clip can also be extracted and saved as an MP3 file. Thus, it is possible for advanced users to take the ambient sound from one clip and use it as the soundtrack for other clips.

The MVS, like an iPod, is not equipped with any speakers. Sound can be heard with the supplied headphones.

3.2.3 Face-to-face transfer

Friends can connect two MVSs together using the supplied USB cable. Movies and clips that are flagged as public can be transfered via USB. When browsing the clips on another MVS, a thumbnail view similar to the local browse view is presented, and the jog wheel can be used in the same way to view thumbnail frames from a given clip.

3.2.4 Digital radio broadcasting

This will be the most technically-challenging aspect of developing the MVS. A wireless networking standard, like 802.11b (at least as I understand it), is focused exclusively on point-to-point data transfer. Point-to-point will not work in this application, since we want to enable one person to broadcast a video in real-time to any number of receivers in the immediate area. Thus, we will likely have to design our own short-range, digital radio standard, possibly using separate frequency bands for each broadcaster in a given vicinity.

As I envision it, one band could be used as a common frequency for announcing broadcast bands. Each broadcaster in a given area would pick a broadcast band at random, announce the band on the common band, and then begin broadcasting a digital MPEG stream on the chosen band.

A receiving MVS would tune to the common band periodically and get an updated list of local broadcast bands. This would enable the instant channel-flipping interface described above.

Since the MPEG data is broadcasted digitally in its original form, it can be saved locally as a clip with no quality loss.

3.2.5 Why not twenty-year-old technology?

The keystone of the MVS is the availability of digital video, which can be copied from user to user indefinitely with no quality loss. In the old world of analog video, something like the MVS was not possible, so instead we saw a rigid, hierarchical structure for relaying video signals from central points to mass audiences (from network headquarters, to regional hubs, to local stations, and finally to individual TV sets). Such structures minimize the number of duplications a program must undergo before it reaches each viewer.

Videos being passed through analog MVSs would be like the circulating bootleg cassette tapes of the 1980s: dubbed generation after generation until all of their content was completely obliterated.

But we *do* have digital video formats, and they have been widely available in the public sphere for more than 10 years. So why has no device like an MVS ever been marketed before? Why is it impossible to connect two iPods together and swap songs (which is something that friends would obviously want to do)? Because companies are scared of handing consumers devices that can limitlessly copy digital media files.

So, media copying and sharing has been pushed onto the Internet, where it simply does not work as well as it could. In fact, even if two friends who want to swap media files have laptops and are sitting next to each other, their best bet is to set up some kind of temporary, two-node Ethernet network. Laptops are supposed to be general-purpose computers (not just media players), and yet their is no way to quickly and easily hook two together and transfer data between them.

For some kinds of media, such as music, it might be reasonable to limit digital copying. After all, as discussed in Appendix A, very few people actually create music, though the vast majority consume music created by others. Thus, whatever digital music is being copied from person-to-person is likely contraband.

With video, however, the story is different. Every person has the ability to author video, so no argument can be made that an MVS-like device would be used primarily for copyright violation.

3.3 Business model

First and foremost, the company would be in the business of manufacturing MVSs and selling them to consumers. Because the MVS uses ten-year-old technology, it will be rather inexpensive to manufacture. I am estimating that a factory could be set up in the U.S. to mass produce MVSs at a retail price point of under $200.

Of course, this is not a sustainable business model: once everyone on the planet owns an MVS, the company producing them will face financial ruin. Other opportunities exist.

For example, many public spaces will invest in fixed MVS broadcast stations. In a shopping mall, for instance, the MVS broadcast station could send out various channels with video presentations about each store in the mall. A museum might have local video broadcasts that augment each exhibit. A stadium might make instant-replays available to everyone in the stands who has an MVS.

The company could engage in service-oriented business to install and maintain these MVS broadcast stations.

In addition, since MVSs become more and more valuable as more people have them, we need to consider services that will make the MVS valuable for early adopters who may not have any friends with MVSs. Thus, some kind of web-based sharing service for MVS videos would be useful, at least in the short term.

A Discussion: The video iPod

There has been a lot of discussion in the press about the possibility of an iPod-like device for video. Several devices like this (often called *personal multimedia players*, or PMPs) are already on the market.

However, an "iPod for video" really misses the mark, because average people interact with video (or at least *could* interact with video) in a different way than they interact with music.

A.1 Music vs. video for "consumers"

Music of any reasonable entertainment value is very difficult to produce—an extraordinary amount of technical skill is necessary. Thus, a very small subset of the population is going to be able to create entertaining music, while the rest of the population is primarily comprised of music consumers.

An iPod is an ideal device for the music landscape, because it is a *one-way* device. Music flows from the creators, down through distribution chains, into iPods, and out through headphones. Music does not flow the other way—into the iPod and out into the world. Indeed, iPods are not equipped with any input capabilities (like microphones or recording features).

On the other hand, entertaining video can potentially be shot by anyone. The only requirement is being in the right place at the right time and knowing which way to point the camera. Witness the fact that some of the most powerful videos of our time were shot by complete amateurs:

- the Kennedy assassination,

- the Rodney King beating,

- the World Trade Center impacts, and

- the recent tsunami land-fall.

The core difference between video and music, perhaps, is that reality does not make for entertaining music. Almost all music is fictional. Producing good *fictional* video is just as difficult as producing good music, but video can be perfectly good without being fictional.

A.2   Video devices must be two-way

Any consumer video device that hopes to be as popular as the iPod is for music must be a two-way device. It must allow consumers to both consume and produce videos. While most consumers understand that producing music is out of reach for them, they *know* that they can produce video—any portable video device that doesn't give them the chance to produce will frustrate them.

Figure 1: Three mock-up views of the MVS (back, front, and bottom).