

# *Benchmarking VoFi*

*A Farpoint Group Technology Note*

*Author: Craig J. Mathias*

Document FPG 2005-068.1  
March 2005



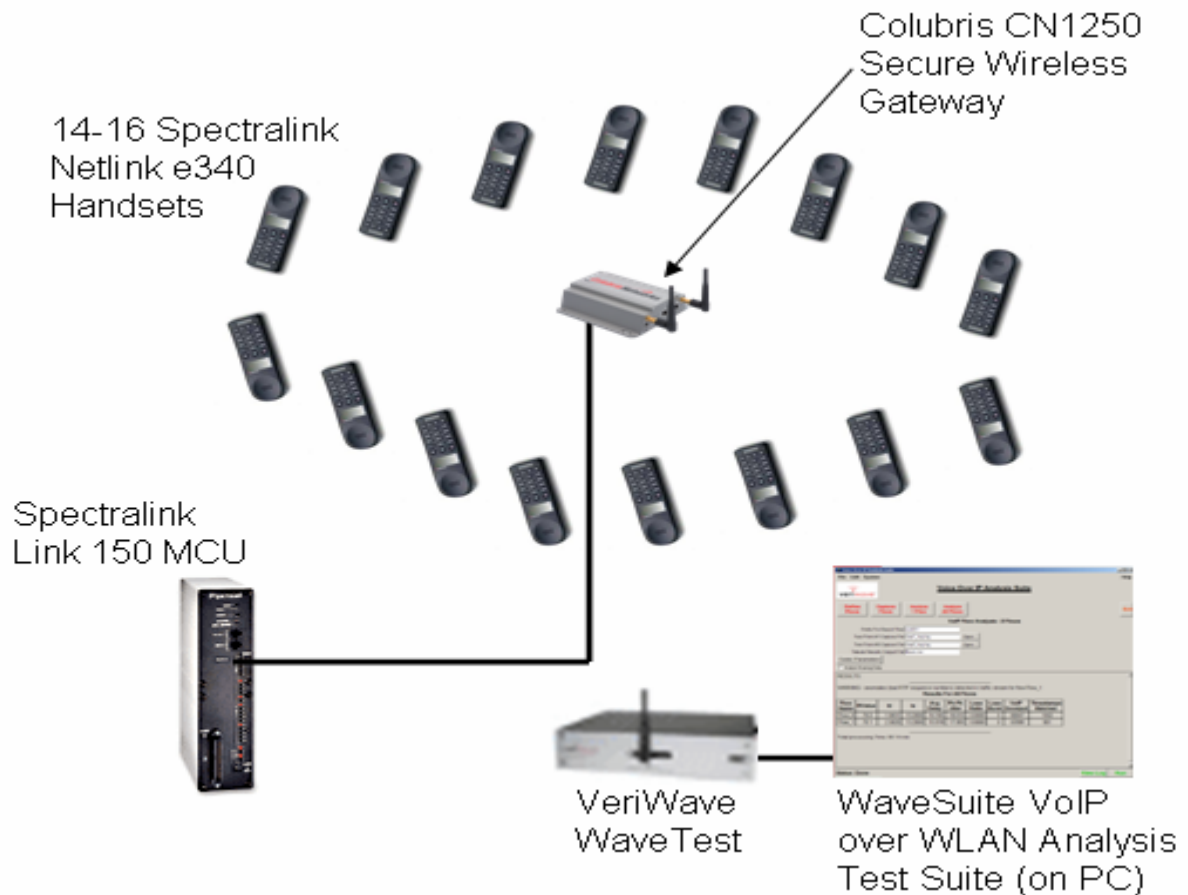
I've received a number of calls from clients recently about VoFi (voice over IP over Wi-Fi), and, specifically, how to test VoFi in production environments. And that's test as in *benchmarking*. You'll recall from my last blurb on this subject [[http://smallbusiness.itworld.com/4384/041129benchwilans/page\\_1.html](http://smallbusiness.itworld.com/4384/041129benchwilans/page_1.html)] that benchmarking wireless systems is tough. There are a large number of variables and elements that otherwise can't be controlled, and comparative tests are therefore to be considered with a good deal of suspicion for this reason. I should have mentioned before that it's possible, in fact, to do laboratory-quality comparative testing by isolating the antennas of the subject systems and feeding them (via RF cabling) into sophisticated test platforms from firms like Azimuth [<http://www.azimuthsystems.com/>], Prosim (Elektrobit) [<http://www.prosim.com/>], and Spirent [[http://www.spirentcom.com/analysis/product\\_product.cfm?PL=23&PS=23&PR=476](http://www.spirentcom.com/analysis/product_product.cfm?PL=23&PS=23&PR=476)]. But these are heavy-duty simulators and analyzers that really demand use by engineers, and are regardless priced well beyond the point of investment by end-user firms. Testing can also be performed inside of wire-mesh "cages", but those are pretty expensive, too. So freespace (over the air, office environment) testing it is.

We recently had the opportunity to do some real-world testing of VoFi handsets. The idea, of course, was to see what kind of performance these products could yield, and the results in this case would relate not to throughput, but rather *voice quality*. Voice is, after all, a medium produced and consumed in real time, with humans on both ends of the connection, as opposed to computers sending bits back and forth. There are a number of ways to determine voice quality, the most common being the Mean Opinion Score (MOS) [<http://www.techabulary.com/m/mos.html>] where respondents are asked to rate the quality of a call (how it sounds) on a 1 to 5 scale. Anything above 4 is considered excellent, but such testing is in and of itself quite subjective.

So, what's needed is something a little more analytical, and we found such a metric in the form of a *Transmission Rating Factor*, usually called *R values*. These are defined by the International Telecommunications Union (ITU) in their standard G.109 [<http://www.kaynam.com/Technology/Docs/ITU-T/G109.doc>], and they can be determined based on a computational/analytical process rather than humans with phones. An R-value of 80 indicates that "all users are satisfied", and scores in the 70s generally correlate to the quality of good cellular voice service. VoIP results can be all over the map, as a function of packet loss. There are, by the way, much more sophisticated ways of evaluating call quality, but they're also much more expensive to implement than the R-value technique.

So, what we set out to do was to set up a properly-instrumented test configuration and run what amounts to a voice benchmark. Other publications have attempted this in the past; see this article in *Network World* [<http://www.nwfusion.com/reviews/2005/011005rev.html>] as an example. In fact, we decided to use a test configuration essentially identical to that used by *Network World*. A diagram of this set-up can be seen in Figure 1, and it consists of the following key components:

- *VoFi handsets and a telephony gateway* – The handsets used were Spectralink [<http://www.spectralink.com>] model e340, along with a Spectralink Link 150 Master



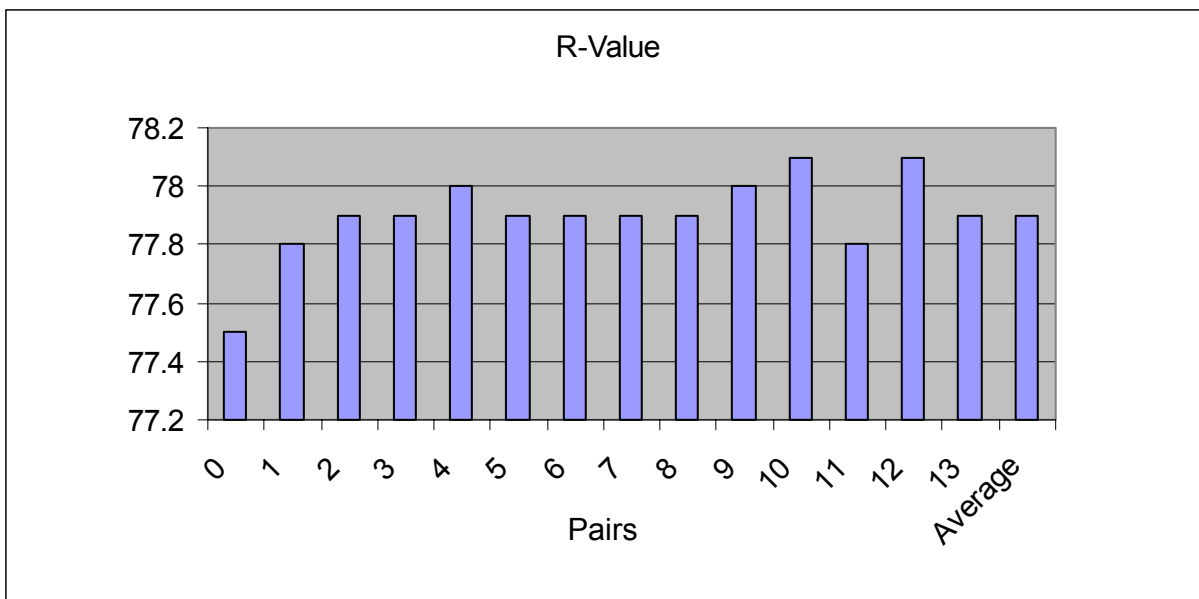
**Figure 1:** The component configuration used in the benchmark tests. *Source:* Farpoint Group.

Control Unit (MCU) gateway. Note that while this is indeed VoFi equipment, it is not compatible with the emerging *Session Initiation Protocol (SIP)* [<http://www.ietf.org/html.charters/sip-charter.html>]. SIP is expected to be the basis of VoFi when it advances into the mass market. The products we used are based on the popular and established H.323 standard, common in many VoIP applications. We used up to 16 (eight pairs) of handsets – any given handset was connected (via a voice call) to one (and only one) other during the test. We didn't attempt to speak into the handsets during the test, so all they communicated was the ambient noise in the room, but that's OK – the amount of traffic in the Spectralink system is always constant when the phone is off-hook. The handsets were arranged in a rough circle about one meter from the Colubris AP (below), the same configuration as was used by *Network World*.

- *Wireless LAN equipment* – The equipment chosen in this case was a Colubris CN1250 gateway [<http://www.colubris.com/Content.aspx?id=245>]. This is the same AP as was used in the *Network World* test, but had improved firmware. Colubris has shipped more than 60,000 APs over the past few years, and has a strong presence in public-access applications. We expect these hotspots to offer voice service once appropriate clients become more common, but you can try using a service like Skype [<http://www.skype.com/>], for example, over a hot-spot network, or your own WLAN, today.

- *Test equipment* – We chose in this case the WaveTest hardware and PC Manager software from VeriWave [<http://www.veriwave.com/>]. The WaveTest is capable of gathering and analyzing a wide variety of WLAN data, and also has the ability to compute R-values via software from VeriWave. We also used the Veriwave equipment to generate interfering traffic in an attempt to degrade overall voice performance; more on this below.

Once all of the equipment was connected, we ran tests using the VeriWave for a period of 30 seconds. The results reported in Figure 2 show the performance measurements results from seven simultaneous handset pairs (the same configuration as were used in the *Network World*

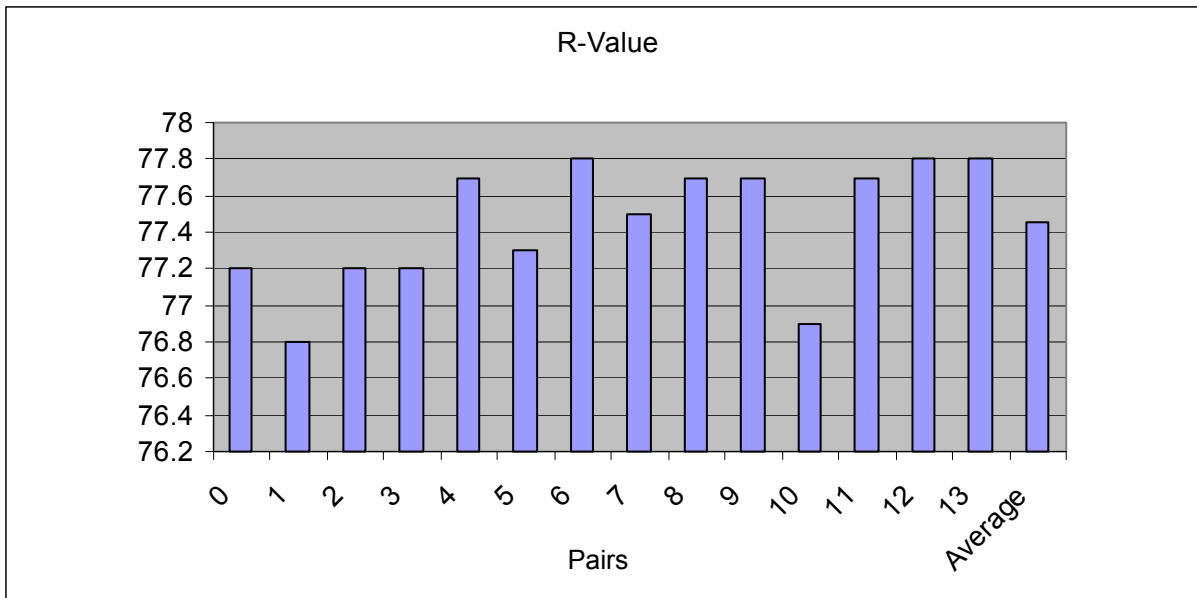


**Figure 2:** Benchmark results for 14 handsets (seven pairs), without interference. *Source:* Farpoint Group.

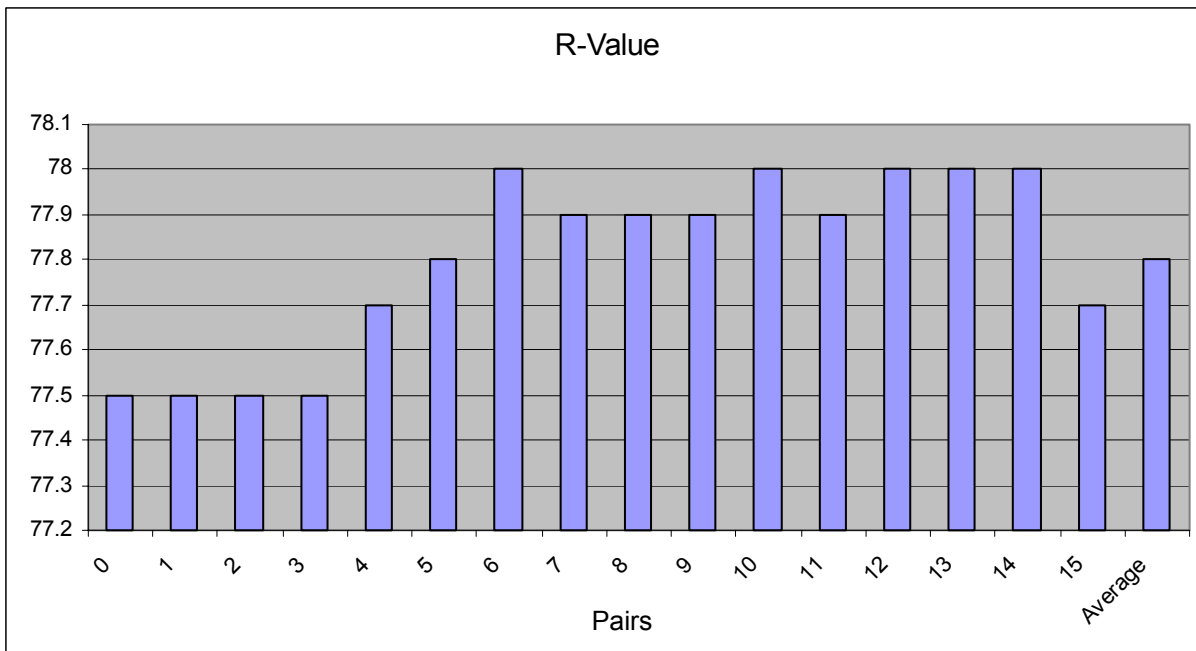
tests). As you can see, the results were excellent and could be expected to yield much more than acceptable voice quality on all pairs – essentially the same as a good cellular connection and many land lines. We then added data traffic at an effective one Mbps rate on the same channel as the handsets, in an attempt to simulate interfering traffic that might be present in a production environment. The *Network World* test also used this same technique. As can be seen in Figure 3, this traffic had no appreciable effect on the benchmark results, apart from slightly greater variance in the performance of individual pairs. Nonetheless, Farpoint Group always recommends assigning voice traffic to its own channel, and not sharing this channel with non-voice users.

We then moved to eight pairs of handsets just for fun. Again, the results were excellent (see Figure 4), with no degradation in the R-value results. While we were mostly looking to just gain experience in VoFi testing, our conclusion is that both the Spectralink and Colubris equipment performed admirably and would make an excellent base for a production VoFi deployment.

The next step will be to test roaming – in our tests the handsets were stationary. The problem in this case is reproducing, exactly, the timing involved and the position of all handsets between benchmark runs with different equipment. We’re going to be experimenting with robots [see <http://www.evolution.com/er1/>] for this purpose; I just bought the first one and I’ll let you know how it works after I get the time to assemble it! The *Network World* roaming test was just too im-



**Figure 3:** Benchmark results for 14 handsets (seven pairs), with interference. *Source:* Farpoint Group.



**Figure 4:** Benchmark results for 16 handsets (eight pairs), with interference. *Source:* Farpoint Group.

precise for my taste, and was really more of a failover test, not indicating all that much, and certainly providing no basis for comparative performance analysis. This lack of precision is a common problem in wireless benchmarks, and a core reason I try to keep tests of this type simple and straightforward.

In conclusion, benchmarking is still fraught with the possibility for error, ambiguity, and misinterpretation of the results. Note that we didn't consider the nature of the environment (interference, multipath, or other artifacts) in the testing described above; it was assumed that any interference would simply result in lower net performance for all handsets. Anyway, the handsets were only about a meter from the access point, so I'm assuming the radio issues were negligible for this test. Regardless, the more experience I gain with wireless benchmarks, the more ideas I have for improving how we should do this in the future. And given the emphasis enterprises place on both comparative and absolute performance, I'm pretty sure I'm going to have the opportunity to try them out soon.



7 Whippoorwill Lane  
Ashland MA 01721  
508-881-6467  
[www.farpointgroup.com](http://www.farpointgroup.com)  
[info@farpointgroup.com](mailto:info@farpointgroup.com)

The information and analysis contained in this document are based upon publicly-available information sources and are believed to be correct as of the date of publication. Farpoint Group assumes no liability for any inaccuracies which may be present herein. Revisions to this document may be issued, without notice, from time to time.

**Copyright 2005 — All rights reserved**

Permission to reproduce and distribute this document is granted provided this copyright notice is included and no modifications are made to the original.