

Millisecond precision psychological research in a world of commodity computers: New hardware, new problems?

RICHARD R. PLANT AND GARRY TURNER
University of York, York, England

Since the publication of Plant, Hammond, and Turner (2004), which highlighted a pressing need for researchers to pay more attention to sources of error in computer-based experiments, the landscape has undoubtedly changed, but not necessarily for the better. Readily available hardware has improved in terms of raw speed; multicore processors abound; graphics cards now have hundreds of megabytes of RAM; main memory is measured in gigabytes; drive space is measured in terabytes; ever larger thin film transistor displays capable of single-digit response times, together with newer Digital Light Processing multimedia projectors, enable much greater graphic complexity; and new 64-bit operating systems, such as Microsoft Vista, are now commonplace. However, have millisecond-accurate presentation and response timing improved, and will they ever be available in commodity computers and peripherals? In the present article, we used a Black Box ToolKit to measure the variability in timing characteristics of hardware used commonly in psychological research.

A common misconception persists that the terms *millisecond accuracy* and *millisecond precision* are interchangeable. Hardware and software vendors often report that their products are “millisecond precise.” However, vendors can guarantee only that measurements are presented with a given timing resolution, or, in effect, with a given granularity, which, in most cases, is in units of milliseconds. But, what do *accuracy* and *precision* mean?

In the fields of science, engineering, industry and statistics, accuracy is the degree of conformity of a measured or calculated quantity to its actual (true) value. Accuracy is closely related to precision, also called reproducibility or repeatability, the degree to which further measurements or calculations show the same or similar results. The results of calculations or a measurement can be accurate but not precise; precise but not accurate; neither; or both. A result is called valid if it is both accurate and precise.

Accuracy vs. precision, “the target analogy.” Accuracy is the degree of veracity while precision is the degree of reproducibility. In this analogy [illustrated in Figure B1], repeated measurements are compared to arrows that are fired at a target. Accuracy describes the closeness of arrows to the bullseye at the target center. Arrows that strike closer to the bullseye are considered more accurate. The closer a system’s measurements to the accepted value, the more accurate the system is considered to be.

To continue the analogy, if a large number of arrows are fired, precision would be the size of the arrow cluster. When all arrows are grouped tightly together, the cluster is considered precise since they all struck close to the same spot, if not necessarily near the bullseye. The measurements are precise, though not necessarily accurate.

—“Accuracy versus precision, the target analogy,”
Wikipedia (2008)

To give a more relevant example, if a stopwatch is held behind one’s back, it could be considered millisecond precise if it ticks regularly once every millisecond; that is, results are displayed in units of a millisecond. However, if you were timing a race, you would need to ensure that you looked at the watch just as the competitors crossed the line. This exemplifies accuracy.

Unfortunately, when conducting computer-based studies, researchers cannot guarantee that timings are either precise or accurate in every case. Vendors of experiment generator packages cannot guarantee that stimulus presentation and response timings are both precise and accurate on all possible combinations of hardware, operating system, scripts, and third-party hardware. Researchers should bear in mind the following formula regarding how various factors can affect timings:

$$\text{True time} = \text{Measured time} - (\text{Error due to paradigm features} + \text{Equipment error})$$

R. R. Plant, r.plant@psych.york.ac.uk

Here, the expression (Error due to paradigm features + Equipment error) encompasses the following:

1. Hardware timing features (e.g., sound card start-up latencies, response device latencies, presentation device latencies);
2. device driver issues and interactions (e.g., drivers that do not work according to the manufacturer's specifications, sluggish drivers, drivers that report incorrect timings);
3. script errors (e.g., poor coding, incorrect settings, human error);
4. effects due to operating system variability (e.g., bugs in the operating system, multicore CPUs, multitasking operating systems);
5. chosen stimulus type (e.g., video presentation, rapid serial visual presentation [RSVP], high-resolution audio);
6. interactions with other software (e.g., antivirus checkers, antispysware packages, hardware "helper" applications);
7. tools used to construct the paradigm (e.g., bugs in experiment generator packages, problems with stimulus preparation packages, such as incorrect video or audio file construction);
8. interactions with other hardware (e.g., synchronicity issues between hardware in paradigms such as fMRI, magnetoencephalography [MEG], eyetracking, psychoacoustics, psychophysics);
9. configuration of settings and levels (e.g., luminance levels, volume settings, crossing thresholds on voice keys);
10. other factors unaccounted for or interactions of the preceding.

As psychologists, we should strive for results that are valid—that is, as precise and as accurate as possible. Any software written by the researcher can also suffer the same inconsistencies, and one should not forget human error on the part of the researcher.

Can one achieve consistently valid, and therefore replicable, results with commodity computers, which, regardless of their apparent power, are ultimately built down to a cost rather than up to a high standard of quality? We suggest that such consistency is possible, but that researchers should exercise extreme caution and that, as Plant, Hammond, and Turner (2004) and Plant, Hammond, and Whitehouse (2002, 2003) advocated, they should validate their presentation and response timings. We accept that authors such as Ulrich and Giray (1989) have suggested that timing inaccuracies may be "overcome" simply by carrying out more trials. The central limit theorem assumes that, although accuracy is present, precision may be lacking. As we have shown, accuracy and precision are in no way guaranteed. Researchers frequently may operate under a different set of conditions than that assumed by Ulrich and Giray.

The central limit theorem also does not solve systemic conditional biases caused either by equipment differences or by paradigm specifics—for example, the finding by Plant et al. (2004) that simply adding a tonal stimulus in a cross-modal priming study resulted in response times that were some 20 msec longer, and with more variability. With

regard to computer mice, Plant et al. (2003) also noted that equipment that may look identical might possess very different timing characteristics. It is not hard to envisage that labs updating equipment and software may find it hard to replicate the results of paradigms from a few years previous. Researchers should bear in mind that "expensive" custom-built experimental machines may often suffer from poor timing characteristics as well. This is not a new problem, and many authors have counseled caution over the years.

It should be noted that, aside from the hardware, application software, and bad device drivers, the operating system can have a huge bearing on how paradigms function. Researchers would do well to be wary of such potential pitfalls. With mainstream experiment generators, such as E-Prime 2.0 from PST, SuperLab from Cedrus, and Inquisit from Millisecond only now beginning to support Microsoft Windows Vista, it would be prudent for researchers to exercise caution. Much debate remains as to whether Microsoft Vista is a suitable core technology for millisecond timing per se, and specifically, whether it is a suitable platform for high-precision audio work (White, 2007), especially in environments in which multitasking is involved. For example, audio prioritization during network activity is now widely recognized as affecting both audio and video (Russovich, 2007).

Also, psychologists often work with third-party hardware and custom peripherals and interfaces, which complicate matters still further. For example, eyetrackers are no longer a rarity in most departments, fMRI and MEG paradigms are now commonplace, and interacting with equipment costing tens of thousands, if not millions, of dollars is not unheard of.

So, in addition to potential conditional biases brought about through hardware and software variations, there is the real prospect of statistically significant effects brought about through use of "limited availability" technology, such as fMRI. That is, one group of participants may be tested by using highly calibrated equipment, whereas another may be tested by using standard commodity PCs in a less well-controlled environment. The question remains as to whether researchers recognize that there may be a concern, and if they do, whether they are willing and able to do something about it.

Unfortunately, due to the delayed release of Microsoft Vista and the fact that common experiment generators are only now beginning to support it, the authors have not had the opportunity to test it thoroughly and fairly. However, to update our articles of 2003 and 2004, which tested computer mice, keyboards, thin-film transistor (TFT) displays, and multimedia projectors (MMPs), it was prudent to investigate whether the timing features of generally available peripherals had changed since those articles were written. This was especially relevant, given that much faster processors and video adapters are now available. Would different TFT panels still introduce variability in stimulus onsets, given newer, faster panels? Would newer Digital Light Processing (DLP) MMPs exhibit the same problems? Would mice still add a variable amount to subjects' response times?

To answer these questions, during 2006, in six tests, we examined a wide range of commonly used commod-

ity peripherals: TFT and CRT displays, MMPs, USB, and PS/2 keyboards and mice, and a variety of speaker systems. Such a range enabled us to test both presentation and response accuracy. Tests were carried out using Windows XP SP2 and E-Prime 1.3 running on a custom-built PC with an AMD Athlon 64 3000+ processor, as detailed in Appendix A. To test the timings of all peripherals, we used a standard Black Box ToolKit (BBTK; www.blackboxtoolkit.com), as detailed in Plant et al. (2004).

We reiterate our goal of stating simply that variations exist among different brands and types of peripherals and that these variations can contribute to spurious or increased effect sizes. The present article addresses only hardware timing features (Point 1 outlined above), and nine or more other sources of timing errors may be present in typical studies. For this article, we used only simple benchmarks. Because paradigms tend to be much more complex and because synchronicity among stimulus types is crucial, the effects of poor timing are likely to be sizeable in real-world studies.

HARDWARE TESTS

TFT Panels

As detailed in Table B1, we tested six TFT panels considered representative of standard commodity TFT panels, for which manufacturers' specifications reported response times of 16–40 msec.

Method. E-Prime was programmed to set pin 1 on the parallel port high for 50 msec and to display simultaneously a 24-bit, 800×600 pixel image at the center of the screen for 120 msec. Pin 1 was set high, so that a BBTK running the Digital Stimulus Capture (DSC) module could detect the leading edge of when the stimulus image should have been displayed, regardless of the panels' onset delay. E-Prime was set to synchronize with the vertical onset of the refresh signal. When E-Prime had displayed the image for 120 msec, a black canvas was shown for 100 msec before the next trial. There were 40 trials in the sequence. A 50-msec assertion of pin 1 was used as a cross-check of E-Prime's internal timing. All panels were driven at 75 Hz via standard Windows settings, giving nine nominal refreshes for a 120-msec image display. As soon as one panel had been tested, the PC was rebooted before the next panel was tested, and so on.

Results. Image-onset delays were calculated by measuring the time from pin 1 on the parallel port going high to the onset of the image being detected by the BBTK opto-detector placed midscreen. Display onset delays range from around 25 to 54 msec. The ordering of these delays largely corresponds to the manufacturer-reported response times. That is, the fastest TFTs had the shortest display onset delays (16 msec), the midrange TFTs clustered around the midrange (20, 25, 30 msec), and the slowest had the longest delay (40 msec). These results are summarized in Figure B2 and Table B2. Ideally, onset delay is as close to 0 msec as possible.

An independent *t* test showed a significant difference between the onset delays of the best- and worst-performing

TFTs, the Dell and MiTAC ($t = 230.55, p < .001$), respectively. A one-way ANOVA found a significant difference among all six TFTs ($F = 19,005.39, p < .001$).

In terms of image duration, which should have been 120 msec, the reverse was true. That is, the TFTs with the slowest response times reported in the manufacturers' specifications terminated soonest, again clustered together in the midrange, and those with the fastest onsets had the longest durations. Display duration times are summarized in Figure B3 and Table B3.

Results from the panels with the fastest manufacturer-reported response times show that they were farthest from the theoretical 120-msec image duration, whereas the slowest had the shortest duration, and the midrange TFTs clustered together around the desired 120 msec. In absolute terms, there was a 28.14-msec difference between the display duration of the best-performing TFT panel and that of the worst. The best was only +0.78 msec from the intended duration of 120 msec; the worst was +20.96 msec.

An independent *t* test showed a significant difference between the display durations of the best- and worst-performing TFTs—namely, the Dell and MiTAC ($t = -699.69, p < .001$). A one-way ANOVA found a significant difference among all six panels tested ($F = 129,971.20, p < .001$).

CRT Monitors

Two CRT monitors were tested, as shown in Table B4.

Method. The methodology and timings were the same as those used in testing TFT panels.

Results. Image onset delays were calculated by measuring the time from pin 1 on the parallel port going high to the onset of the image being detected by the BBTK opto-detector. In both cases, the display onset delay was around +20 msec, as shown in Figure B4 and Table B5. This is exactly one refresh at 75 Hz (13.33 msec) plus half a refresh for the raster to trigger the opto-detector at midscreen (6.66 msec).

An independent *t* test showed no significant difference between the onset delays of the two CRTs ($t = 0.14, p = .89$).

As regards image duration, the LG CRT consistently terminated at 132 msec, whereas the Sony CRT terminated at 143 msec, as shown in Figure B5 and Table B6.

Results show an absolute difference of 11.71 msec between the CRTs. There was only a +11.36-msec difference from the theoretical 120 msec for the LG display and +23.07 msec for the Sony display.

An independent *t* test showed a significant difference between the display durations of the two CRTs ($t = -571.839, p < .001$).

MMPs

Three MMPs were tested, as shown in Table B7.

Method. The methodology and timings were the same as those used in testing TFT panels.

Results. Again, image onset delays were calculated by measuring the time from pin 1 on the parallel port going high to the onset of the image being detected by the BBTK opto-detector. The X328 had a mean onset delay

of +29.78 msec and the L850E of +35.30 msec, whereas the Dell 2300MP had a much more consistent delay of +30.20 msec, as shown in Figure B6 and Table B8.

An independent *t* test showed a significant difference between the onset delays of the best- and worst-performing MMPs—namely, the Dell 2300MP and the Hitachi CPL850E ($t = -6.58, p < .001$). A one-way ANOVA found a significant difference among all three MMPs tested ($F = 25.036, p < .001$).

Mean image durations for the two Hitachi projectors were consistently identical at 121.78 msec, whereas termination for the Dell was 139.75 msec, as shown in Figure B7 and Table B9.

Results for the best-performing Dell projector show an image duration of 139.75 msec, which is +19.75 msec from the intended 120 msec. Both Hitachi projectors have identical characteristics, with a mean duration of +0.78 msec from intended.

An independent *t* test showed a significant difference between the display durations of the best- and worst-performing MMPs ($t = 885.74, p < .001$). A one-way ANOVA found a significant difference among all three MMPs tested ($F = 394,569.1, p < .001$).

Mice

Eleven mice were tested, as detailed in Table B10. These included PS/2, USB, serial, and wireless mice.

Method. E-Prime was programmed to send pin 1 on the parallel port high for 80 msec. When pin 1 went low, the BBTK, running the Digital Stimulus Capture and Response (DSCAR) module, sent an active switch closure message to the left button of a mouse under test. When E-Prime detected the button being pressed (i.e., a simulated response), it made pin 2 of the parallel port go high for 80 msec. There was a 100-msec interval between each of the 40 trials.

Results. Response time delays were calculated by subtracting the time at which pin 2 went high from the time at which pin 1 went low again. All timings were captured and calculated using the BBTK. The best-performing mouse, the Kensington Mouse, consistently added +9.52 msec to RT error, whereas the worst three added around +49 msec and showed much more variability. As expected, a wireless mouse was among the worst performers (Chic Wireless Optical Mouse), adding around +48.96 msec of error. However, the second wireless mouse (Microsoft Standard Wireless Optical Mouse) was more consistent, adding around +28.86 msec of error. Others, such as the Microsoft Wheel Mouse Optical, were remarkably consistent, showing very little variation. These results are summarized in Figure B8 and Table B11.

An independent *t* test showed a significant difference between the response times of the best- and worst-performing mice, the Kensington Mouse and the Chic Wireless Optical Mouse ($t = -49.173, p < .001$). A one-way ANOVA found a significant difference among all 11 tested ($F = 816.68, p < .001$).

Keyboards

Four keyboards were tested, as shown in Table B12. These included PS/2, USB, and wireless models.

Method. The methodology and timings were the same as those used in testing computer mice, although here the space bar was wired to a BBTK Active Switch Closure, to which E-Prime was scripted to respond by sending pin 2 high on the parallel port.

Results. Again, response time delays were calculated by subtracting the time at which pin 2 went high from the time at which pin 1 went low again. The BBTK DSCAR module was used to capture and calculate all timings. The best-performing keyboard (Belkin Classic USB) added around +18.30 msec to response times with a low variability; the worst (DAN 102 Key Keyboard) added a mean of +33.73 msec with a large variability. The Chic Intelligent Wireless Keyboard added a mean of +25.34 msec, again with quite a large variability. These results are summarized in Figure B9 and Table B13.

An independent *t* test showed a significant difference between the response times of the best- and worst-performing keyboards—namely, the Belkin Classic and DAN 102 ($t = -29.29, p < .001$). A one-way ANOVA found a significant difference among all four keyboards tested ($F = 305.88, p < .001$).

Computer Speaker Systems

As shown in Table B14, eight computer speaker systems were tested. These included standard 3.5-mm stereo jack, amplified speakers, and USB-powered/interfaced models. We made a cursory examination of the use of DSP (digital signal processing) effects. All USB-interfaced speakers used Microsoft Windows XP's built-in USB HID (human interface device) drivers. A Creative Labs Sound Blaster Live Value! PCI card with WHQL (Windows Hardware Quality Laboratories) Windows XP drivers was used for all sound production, because this configuration was known to have extremely low start-up latencies when used with E-Prime under Windows XP.

Method. E-Prime was programmed to send pin 1 on the parallel port high for 80 msec and then play a 22.050-Hz, 16-bit, stereo tone lasting 1.00 sec. The BBTK DSC module was used to both monitor pin 1 on the parallel port and record the onset of the sound with a BBTK digital microphone.

Results. Delays introduced by the various speaker systems and technologies were calculated by subtracting the time at which the leading edge of the sound was detected from the time at which pin 1 on the parallel port went low. As expected, the poorest-performing speaker systems were those using the USB interface. The worst added 10.12 msec of start-up time to the sound being heard; the best added as little as 3.21 msec. These results are summarized in Figure B10 and Table B15.

An independent *t* test showed a significant difference between the onset times for the best- and worst-performing systems (DSP aside): the Encore P-905U and the Yamaha YST-M20DSP (no DSP) ($t = 34.36, p < .001$). A one-way ANOVA found a significant difference among all eight systems tested ($F = 522.66, p < .001$).

Some systems either featured hardware DSP effects built into the amplifier or replicated this functionality within software. The best-performing system, the Ya-

maha YST-M20DSP, features a hardware DSP, whereas the worst-performing, the Jazz, processes DSP within its driver software. On the Yamaha speakers, having DSP activated had little effect on sound onset times, whereas, when using the Jazz's own driver (i.e., using the default setting for the driver, as per an initial installation), a large difference was noted, as shown in Table B16. This "generic" setting could not be switched off when Jazz's driver was used, although other DSP effects could be selected.

For the Jazz speakers, an independent *t* test showed a significant difference in onset times between use of the Windows XP HID USB speaker drivers and Jazz's own drivers, which had generic DSP switched on by default ($t = -102.63, p < .001$). The effect on onset delay is illustrated clearly in Figure B11.

DISCUSSION

Most psychologists would agree that, for fast visual stimulus presentation, such as RSVP, CRT monitors are preferable to TFT panels. They often cite the fact that panels have slower response times and that images do not appear when intended. Indeed, our empirical testing would suggest this to be so, with onset delays for the six panels tested falling in a range between mid-20 to mid-50 msec. It is also worth noting that the response times reported in the manufacturers' specifications did correlate somewhat with image onset delay. More unexpected is the effect that this had on actual display duration (i.e., the period during which the image was visible, as compared with the target of 120 msec). Although the best panels were a mere +0.78 msec adrift, the worst were as much as +20.96 msec adrift. It is plausible to imagine a situation in which a high-quality panel was used in a high-precision environment (e.g., fMRI and where a lower specification model was used back in the departmental lab). Often, each location is tied to a particular condition or experimental manipulation (e.g., patients vs. students, respectively). Even with advanced time-auditing features, commercial systems, such as E-Prime, cannot be aware of or allow for performance of display devices with which they operate. Factors, such as luminance varying among panels, can also have a marked effect.

With respect to TFT panels and MMPs, the computer's graphics card indicates to the operating system and application software when a nominal refresh signal occurs, but this indication does not affect when the image appears on the TFT, because the technology involved does not support the ability to display images as a series of frames as CRTs do. The image is updated according to the panel's electronics, independently of the graphics card. The computer cannot detect when the panel draws the image from its own buffering mechanism. The response times reported in the manufacturers' specifications represent the panel's internal redrawing time. Our data show that buffering times vary across manufacturers and that the reported times may be misleading in absolute terms. Schmidt (2001) pointed out that researchers

often are unaware that presentation times vary and that the actual presentation times may differ from what the computer reports. Schmidt used the term *presentation accuracy* to refer to the actual presentation timings when measured using external measures (i.e., image onset, duration, offset).

One way to help ensure that stimulus displays perform as intended is to use an external chronometry device, such as the BBTK, and to adjust timings manually, in order to correct for any hardware variance (e.g., make stimuli appear sooner in the experimental timeline or increase/decrease their duration). In the present study, when all six panels were compared, there was a statistically significant difference between the best- and worst-performing panel, whereas, when traditional CRTs were used, there was no significant difference. Albeit, one CRT was consistently one refresh period slower than the other.

MMPs produced variable results in terms of onset delays and actual image duration. Again, statistically significant differences were found between the best- and worst-performing projectors. There was also a significant difference among the three models tested. One can conclude that TFT panels do not have the same display characteristics as older CRTs and that MMPs are not equivalent to either.

Having dealt with visual presentation devices, our attention was turned to response devices with 11 computer mice examined. Testing revealed a huge spread in the contribution each mouse added to response time. There was a significant difference among all the mice tested, and, as expected, especially between the best- and worst-performing mice. It should be noted that performance could not be predicted on the basis of brand, cost, interface (e.g., USB, PS/2), technology (ball vs. optical), or wireless versus wired.

In much the same way as did mice, keyboards showed a range of variability. Again, there was an unpredictable difference between the best and the worst that proved significant. When considered together, there was a significant difference among all devices.

Computer speakers showed the most promise in terms of onset speed, duration, and consistency. All speakers responded with onset of less than 10 msec, and durations were within 1–2 msec of those intended. It is accepted that the Creative Labs SoundBlaster Live! PCI card will have added a small, but consistent, start-up latency. Although differences were small in absolute terms, they were statistically significant. This card is known to have very low start-up latency and is not representative of all sound cards, which can have longer and more variable latencies.

A tantalizing insight into the effect of driver software was given when the native drivers were loaded onto the system for the Jazz USB speakers. The mean onset delay went to 37 msec from the 9.68 msec seen with Windows XP's native HID drivers for USB speakers, with no change in the hardware or other configuration. As mentioned in the introduction, some observers remain unconvinced that Mi-

Microsoft Vista will ever be a suitable operating system for high-precision work. Here, for example, we can see a degradation of some tens of milliseconds as a result of using one driver over another with the same hardware. It should be remembered that software, regardless of whether it be the operating system itself or a poorly written driver, can have an undesirable, and indeed unknown, negative effect on timing. This phenomenon is certainly not new (e.g., Krantz, 2000), and it suggests that both hardware and software can introduce additional variability in experimental timing. In an effort to reduce costs of commodity computers, manufacturers often emulate hardware using software drivers. The Yamaha speakers with switchable hardware DSP effects, for example, showed no marked increase in latencies, regardless of whether DSP was switched on.

In the present article, we have examined the contribution that commodity hardware can make toward timing inaccuracies. This was done by using very simple E-Prime scripts, which were geared toward highlighting differences between hardware rather than being representative of paradigms run by researchers. In the research environment, one should consider many factors that might affect timing, such as the contributions made by the operating system, hardware drivers and their supporting software, the software package in which the paradigm runs, the computer mainboard, the graphics card, and the sound card. One should also account for human error (e.g., scripting errors, incorrect timings in complex code, stimulus preparation errors). In short, a multitude of factors can affect timing accuracy. Often, paradigms present stimuli in such a way that problems may not be obvious to the researcher.

Multitasking operating systems make high-precision experimental control harder rather than easier. A faster multicore machine does not guarantee better timing. This applies to Microsoft Windows just as it does to Macintosh and Linux operating systems. Timing inaccuracy is not a new phenomenon, with Beringer (1992) first highlighting a specific issue with regard to the response timing of computer mice. Ulrich and Giray's (1989) advice on carrying out more trials to overcome timing inconsistencies may no longer be as applicable to today's paradigms and complex experimental setups as it was then.

A researcher reporting results in units of milliseconds might consider independent third-party validation as a way of ensuring tighter experimental control and enhancing credibility in one's field of expertise. We propose that this should be done with an unmodified paradigm running in situ on the experimenter's own hardware. It is not hard to envisage a time when journals request some assurance

that presentation and response timings are as accurate and precise as possible and that authors are willing to state what the confidence limits are for a given study. We feel that researchers in other scientific disciplines would take a dim view of the lack of experimental candor shown in the majority of academic articles published within psychology journals without reference to experimental controls or with statements of precision and accuracy. It is important to note that the fact that hardware and software produce answers that "look accurate" does not mean that those answers are valid.

AUTHOR NOTE

This research was first presented during a symposium at the Society for Computers in Psychology (SCiP) conference, Long Beach, CA, November 15, 2007. We thank William Schmidt for his constructive comments in the redrafting of the present article. Address correspondence to R. R. Plant, Department of Psychology, University of York, Heslington Road, York YO10 5DD, England (e-mail: r.plant@psych.york.ac.uk).

REFERENCES

- ACCURACY VERSUS PRECISION, THE TARGET ANALOGY (n.d.). Retrieved March 31, 2008, from Wikipedia: http://en.wikipedia.org/wiki/accuracy_and_precision#accuracy_versus_precision.3B_the_target_analogy.
- BERINGER, J. (1992). Timing accuracy of mouse response registration on the IBM microcomputer family. *Behavior Research Methods, Instruments, & Computers*, **24**, 486-490.
- KRANTZ, J. H. (2000). Tell me, what did you see? The stimulus on computers. *Behavior Research Methods, Instruments, & Computers*, **32**, 221-229.
- PLANT, R. R., HAMMOND, N., & TURNER, G. (2004). Self-validating presentation and response timing in cognitive paradigms: How and why? *Behavior Research Methods, Instruments, & Computers*, **36**, 291-303.
- PLANT, R. R., HAMMOND, N., & WHITEHOUSE, T. (2002). Toward an experimental timing standards lab: Benchmarking precision in the real world. *Behavior Research Methods, Instruments, & Computers*, **34**, 218-226.
- PLANT, R. R., HAMMOND, N., & WHITEHOUSE, T. (2003). How choice of mouse may affect response timing in psychological studies. *Behavior Research Methods, Instruments, & Computers*, **35**, 276-284.
- RUSSINOVICH, M. (2007, August 27). Vista Multimedia playback and network throughput. Article posted to <http://blogs.technet.com/markrussinovich/archive/2007/08/27/1833290.aspx>
- SCHMIDT, W. C. (2001). Presentation accuracy of Web animation methods. *Behavior Research Methods, Instruments, & Computers*, **33**, 187-200.
- ULRICH, R., & GIRAY, M. (1989). Time resolution of clocks: Effects on reaction time measurement: Good news for bad clocks. *British Journal of Mathematical & Statistical Psychology*, **42**, 1-12.
- WHITE, N. (2007, October 29). An overview of Windows sound and music "glitching" issues. Article posted to <http://windowsvistablog.com/blogs/windowsvista/archive/2007/10/29/an-overview-of-windows-sound-and-music-glitching-issues.aspx>.

APPENDIX A Test PC Specifications

Component	Description
Processor	Model: 1x AMD Athlon 64 Processor 3000+ Speed: 2.07 GHz Internal data cache: 64 K, synchronous, write-back, 2-way set-associative, 64-byte line Internal instruction cache: 64 K, synchronous, write-back, 2-way set-associative, 64-byte line L2 on-board cache: 512 K, ECC, synchronous, write-back, 16-way set-associative, 64-byte line

APPENDIX A (Continued)

Component	Description
System	Mainboard: NF-CK804 Chipset: nVidia nForce4 and Ultra MCP System BIOS: Phoenix Technologies, LTD 6.00 PG Total memory: 1-GB DDR-SDRAM
Chipset	Model: Advanced Micro Devices (AMD) Athlon 64/Opteron HyperTransport Technology Configuration Front-side bus speed: 2× 920 MHz (1840-MHz data rate) Total memory: 1-GB DDR-SDRAM
Video system	Adapter: Radeon X550 PCIe x16 Chipset: ATI Radeon Graphics Processor (0x5B63) RAMDAC: Internal DAC (400 MHz) Total memory: 256 MB (256-MB video) (246-MB system) Texture memory: 507 MB
Sound adapter	Model: Creative Labs CT4830 SBLive! Value Revision: A8 Type: AC '97
USB controller	Model: Foxconn International Inc nForce4 USB 2.0 Controller Revision: K4 Version: 2.00 Specification: 1.00 Interface: EHCI Channels: 10 Companion controllers: 1 Supported speeds: low (1.5 Mbps), full (12 Mbps), high (480 Mbps) Addressing support: 32 bit Legacy emulation enabled: No
Operating system	Windows system: Microsoft Windows XP (2002) Professional 5.01.2600 (Service Pack 2) Platform compliance: Win32 x86
Network services	Adapter: D-Link DGE-528T Gigabit Ethernet adapter

Note—Test PC specification was obtained using SiSoft Sandra Engineer 2007.

APPENDIX B

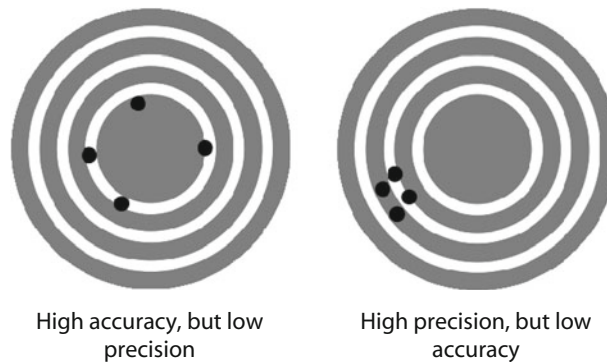


Figure B1. Accuracy versus precision, “the target analogy.”

**Table B1
TFT Panels Tested and Manufacturer-Reported
Response Times (RTs, in Milliseconds)**

Model	Size (in.)	Native Resolution (pixels)	RT
Acer AL718	17	1,280 × 1,024	20
Viglen EZX15F	15	1,024 × 768	30
AOC LM720A	17	1,280 × 1,024	25
MiTAC MT-15LXA	15	1,024 × 768	40
Dell E156FPf	15	1,024 × 768	16
LG Flatron L1715SM	17	1,280 × 1,024	16

Note—All panels used a 15-pin D-SUB interface.

APPENDIX B (Continued)

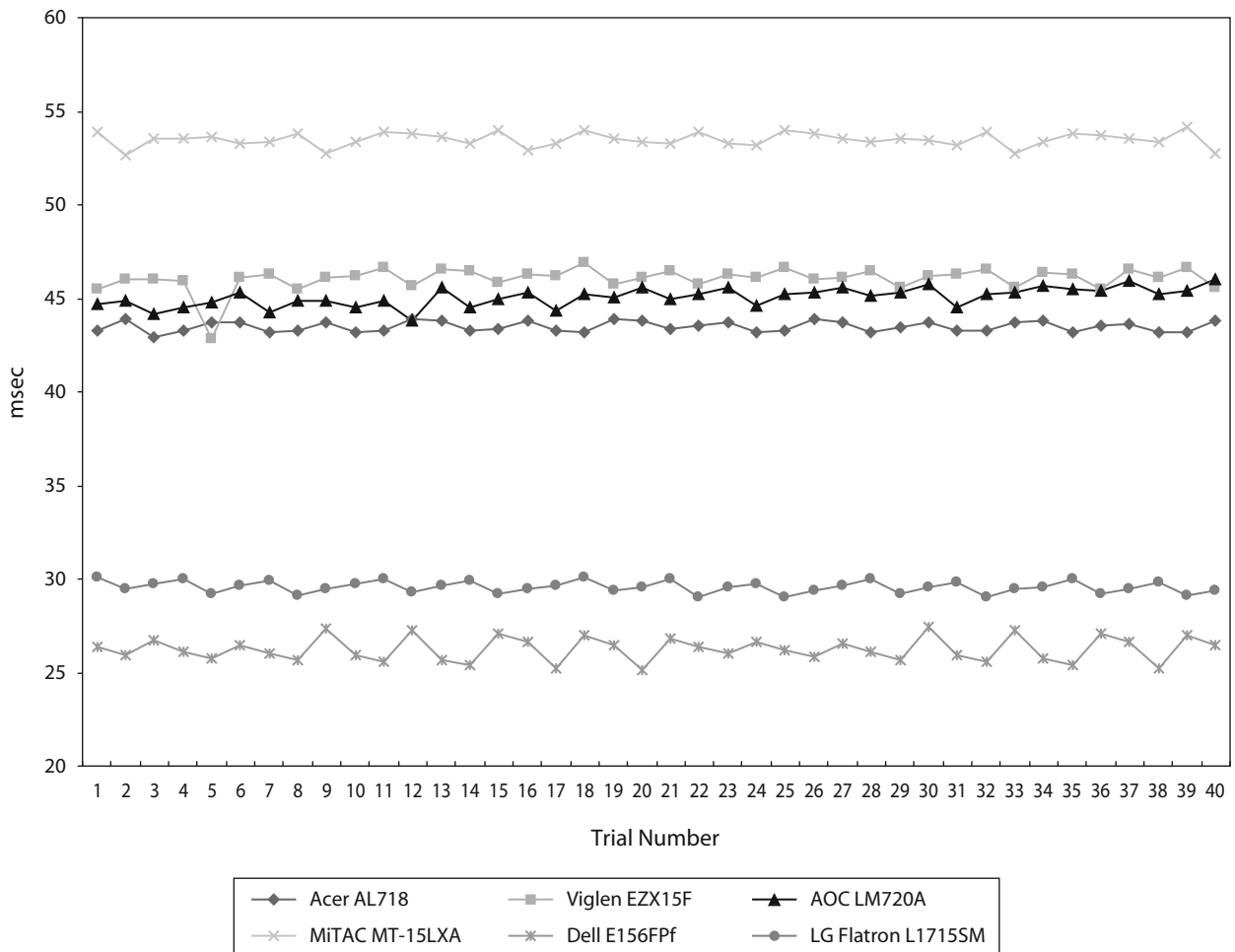


Figure B2. Summary of image onset delays for thin film transistor (TFTs) (target = 0 msec).

Table B2
Summary of Image Onset Delays for Thin Film Transistor (TFTs, in Milliseconds)

Model	Image Onset Delay	
	<i>M</i>	<i>SD</i>
Acer AL718	43.48	0.28
Viglen EZX15F	46.04	0.63
AOC LM720A	45.07	0.50
MiTAC MT-15LXA	53.48	0.38
Dell E156FPf	26.24	0.65
LG Flatron L1715SM	29.56	0.32

APPENDIX B (Continued)

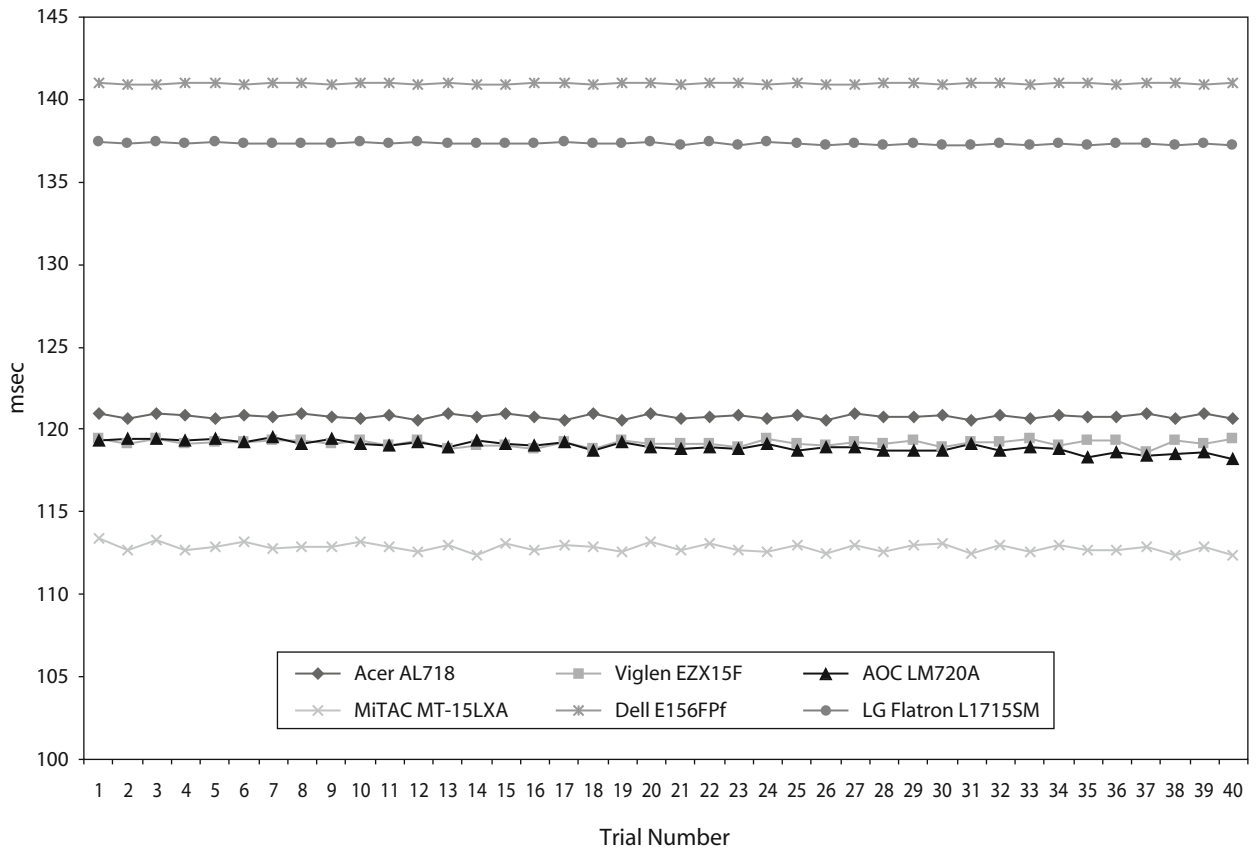


Figure B3. Summary of image durations for TFTs (target = 120 msec).

Table B3
Summary of Image Durations
for TFTs (in Milliseconds)

Model	Image Duration	
	<i>M</i>	<i>SD</i>
Acer AL718	120.78	0.13
Viglen EZX15F	119.17	0.20
AOC LM720A	118.97	0.33
MiTAC MT-15LXA	112.82	0.25
Dell E156FPf	140.96	0.03
LG Flatron L1715SM	137.33	0.06

Table B4
Cathode Ray Tube (CRT) Monitors Tested

Model	Size (in.)	Interface Used
LG Studioworks 775E	17	15-pin D-SUB
Sony Trinitron 110ES	15	15-pin D-SUB

APPENDIX B (Continued)

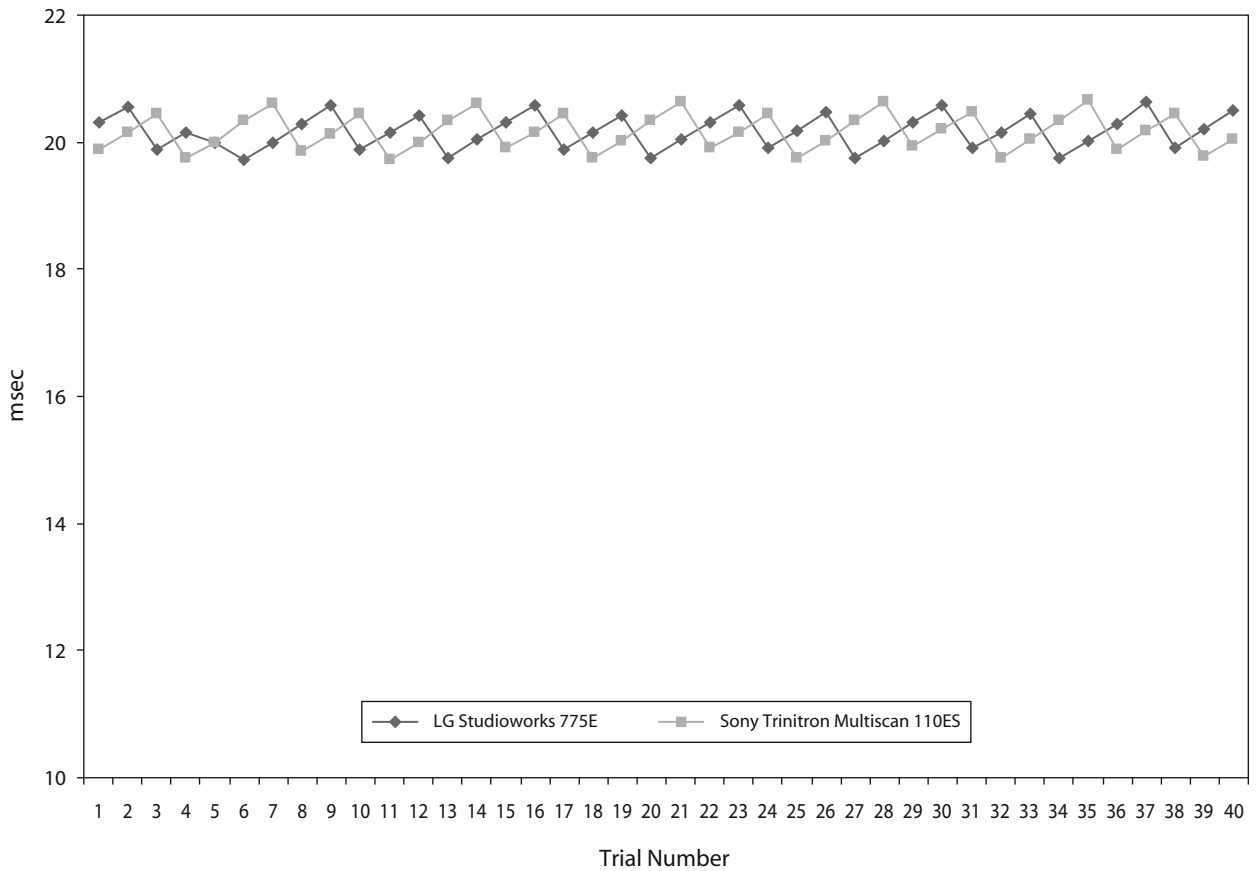


Figure B4. Summary of image onset delays for CRTs (target = 0 msec).

Table B5
Summary of Image Onset Delays
for CRTs (in Milliseconds)

Model	Image Onset Delay	
	<i>M</i>	<i>SD</i>
LG Studioworks 775E	20.17	0.28
Sony Trinitron 110ES	20.16	0.29

APPENDIX B (Continued)

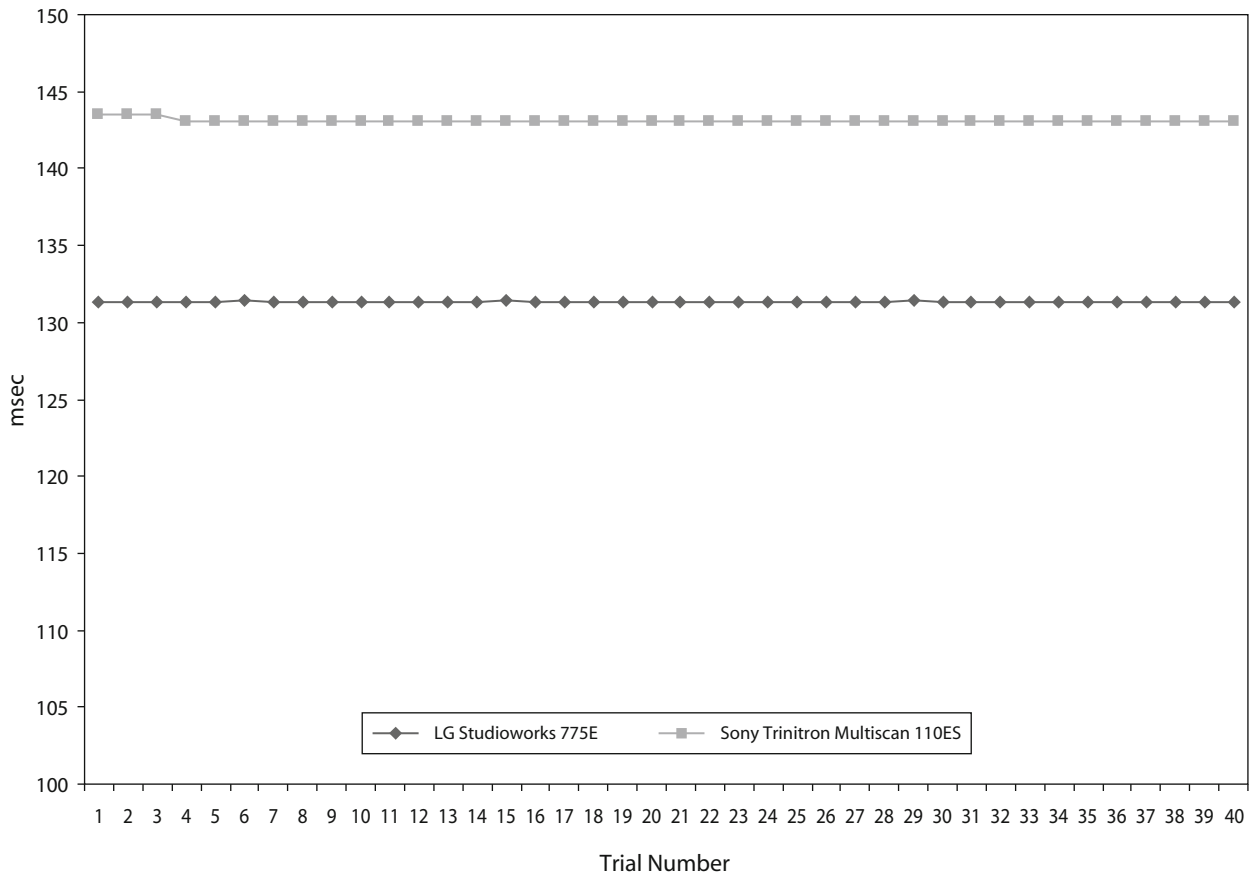


Figure B5. Summary of image durations for CRTs (target = 120 msec).

Table B6
Summary of Image Durations for CRTs

Model	Image Duration	
	<i>M</i>	<i>SD</i>
LG Studioworks 775E	131.36	0.01
Sony Trinitron 110ES	143.07	0.13

Table B7
Multimedia Projectors (MMPs) Tested

Model	Native Resolution (pixels)	Interface Used	Technology
Dell 2300MP	1,024 × 768	15-pin D-SUB	DLP
Hitachi CPL850E	1,024 × 768	15-pin D-SUB	LCD
Hitachi CP-X328	1,024 × 768	15-pin D-SUB	LCD

APPENDIX B (Continued)

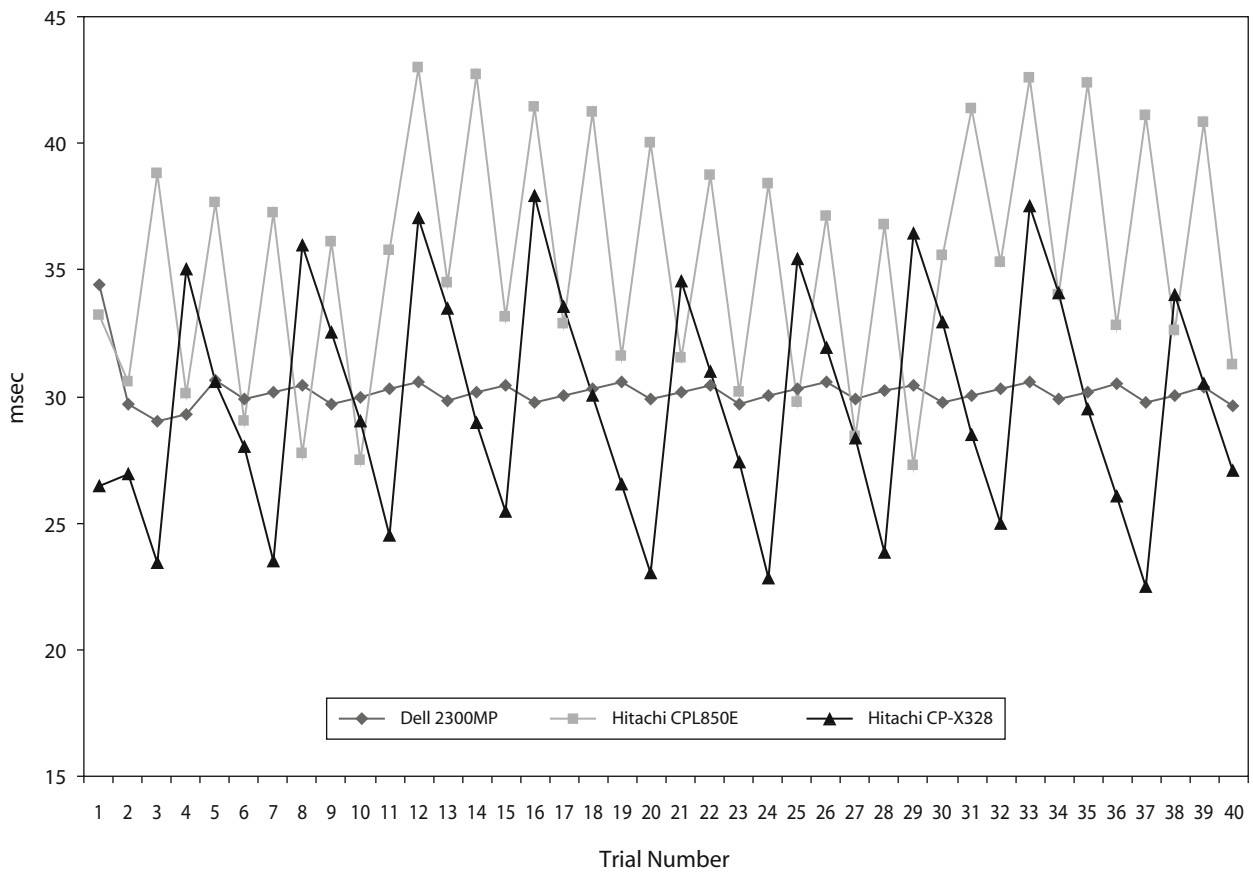


Figure B6. Summary of image onset delays (in milliseconds, target = 0 msec) for multimedia projectors (MMPs).

Table B8
Summary of Image Onset Delays
for Multimedia Projectors
(MMPs, in Milliseconds)

Model	Image Onset Delay	
	<i>M</i>	<i>SD</i>
Dell 2300MP	30.20	0.78
Hitachi CPL850E	35.30	4.84
Hitachi CP-X328	29.78	4.60

APPENDIX B (Continued)

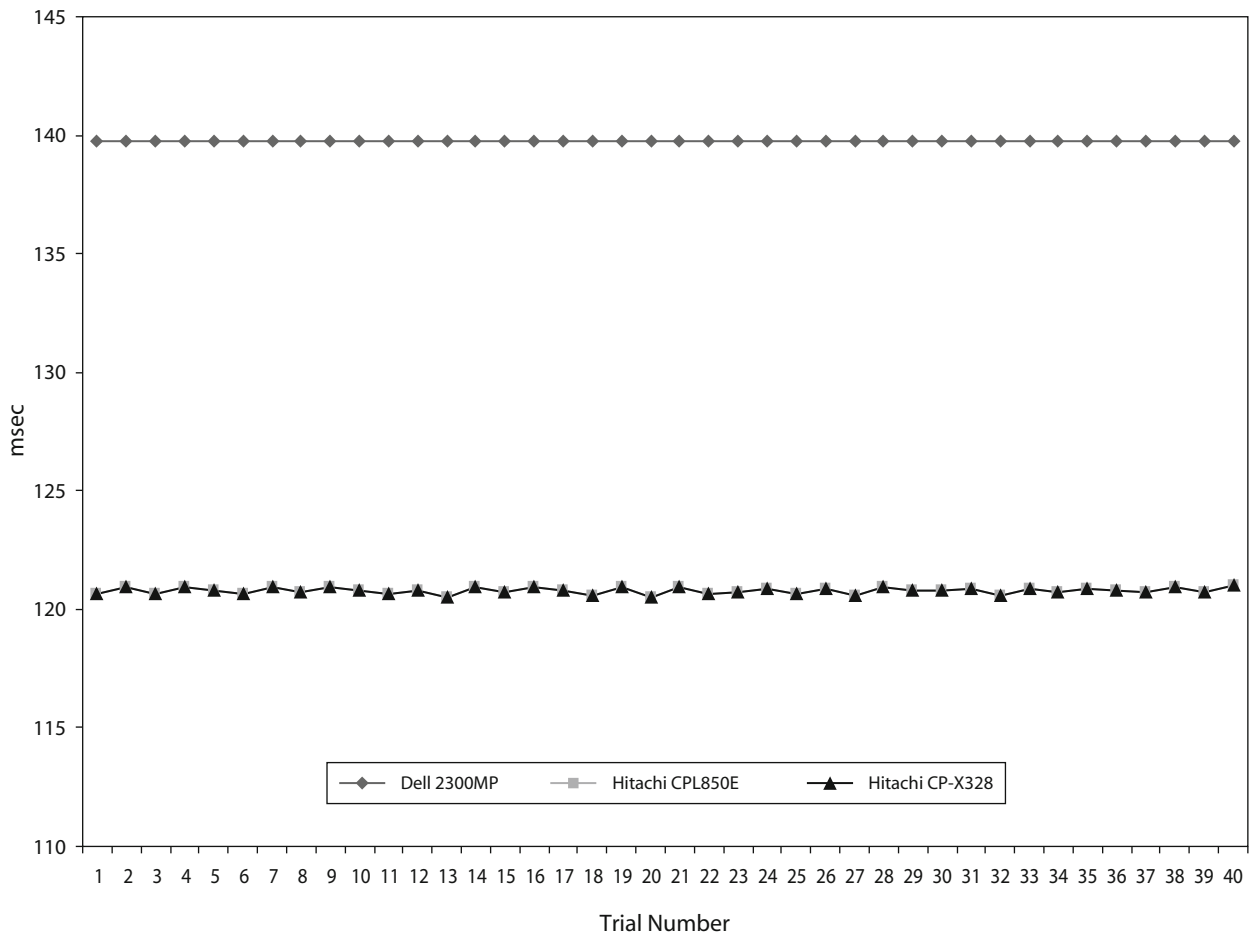


Figure B7. Summary of image durations (in Milliseconds, target = 120 msec) for multimedia projectors (MMPs).

Table B9
Summary of Image Durations
for Multimedia Projectors
(MMPs, in Milliseconds)

Model	Image Duration	
	<i>M</i>	<i>SD</i>
Dell 2300MP	139.75	0.01
Hitachi CPL850E	120.78	0.13
Hitachi CP-X328	120.78	0.13

Table B10
Computer Mice Tested

Model	Interface
Alice Mouse FCC-ID_IOWCM-PS2C	PS/2
Chic Wireless Optical Mouse	USB
Kensington Mouse	PS/2
Microsoft Standard Wireless Optical Mouse	USB
Microsoft Wheel Mouse	USB
Trust Ami Mouse Dual Scroll 11933	Serial
Chic Wireless Optical Mouse	PS/2
IntelliMouse 72113 Mouse-in-a-Box Optical	USB
Mouse 1300	PS/2
Microsoft Wheel Mouse Optical	PS/2
Trust Ami Mouse Dual Scroll	PS/2

APPENDIX B (Continued)

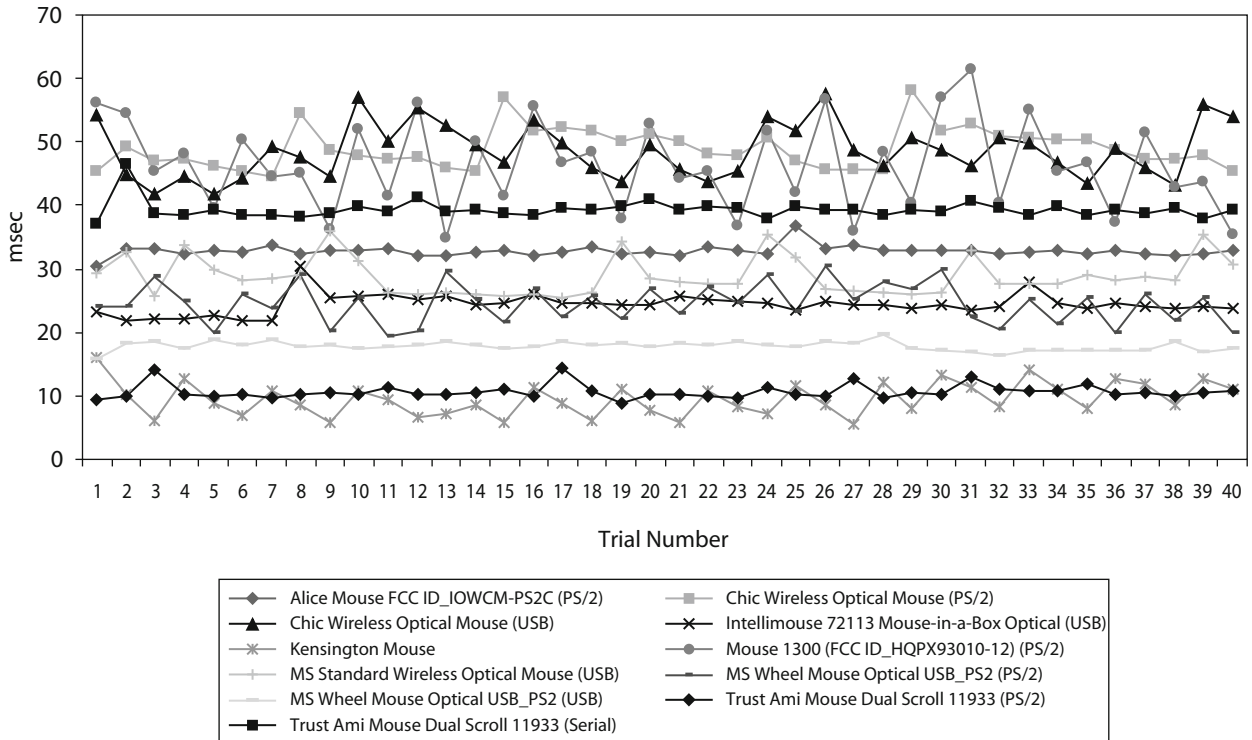


Figure B8. Summary of response time delays for mice tested (target = 0 msec).

Table B11
Summary of Response Time Delays for Mice Tested

Model (Interface)	Response Time	
	<i>M</i>	<i>SD</i>
Alice Mouse FCC ID_IOWCM-PS2C (PS/2)	32.79	0.85
Chic Wireless Optical Mouse (PS/2)	48.96	3.19
Chic Wireless Optical Mouse (USB)	48.58	4.28
IntelliMouse 72113 Mouse-in-a-Box Optical (USB)	24.44	1.59
Kensington Mouse	9.52	2.63
Mouse 1300 (FCC ID_HQPX93010-12) (PS/2)	46.39	7.07
Microsoft Standard Wireless Optical Mouse (USB)	28.86	3.02
Microsoft Wheel Mouse Optical USB_PS2 (PS/2)	24.52	3.14
Microsoft Wheel Mouse Optical USB_PS2 (USB)	17.80	0.70
Trust Ami Mouse Dual Scroll 11933 (PS/2)	10.68	1.16
Trust Ami Mouse Dual Scroll 11933 (Serial)	39.31	1.44

Table B12
Keyboards Tested

Model	Interface
Belkin Classic Keyboard	USB
DAN 102 Key Keyboard	PS/2
Chic Intelligent Wireless Keyboard	PS/2
Smart Keyboard	PS/2

APPENDIX B (Continued)

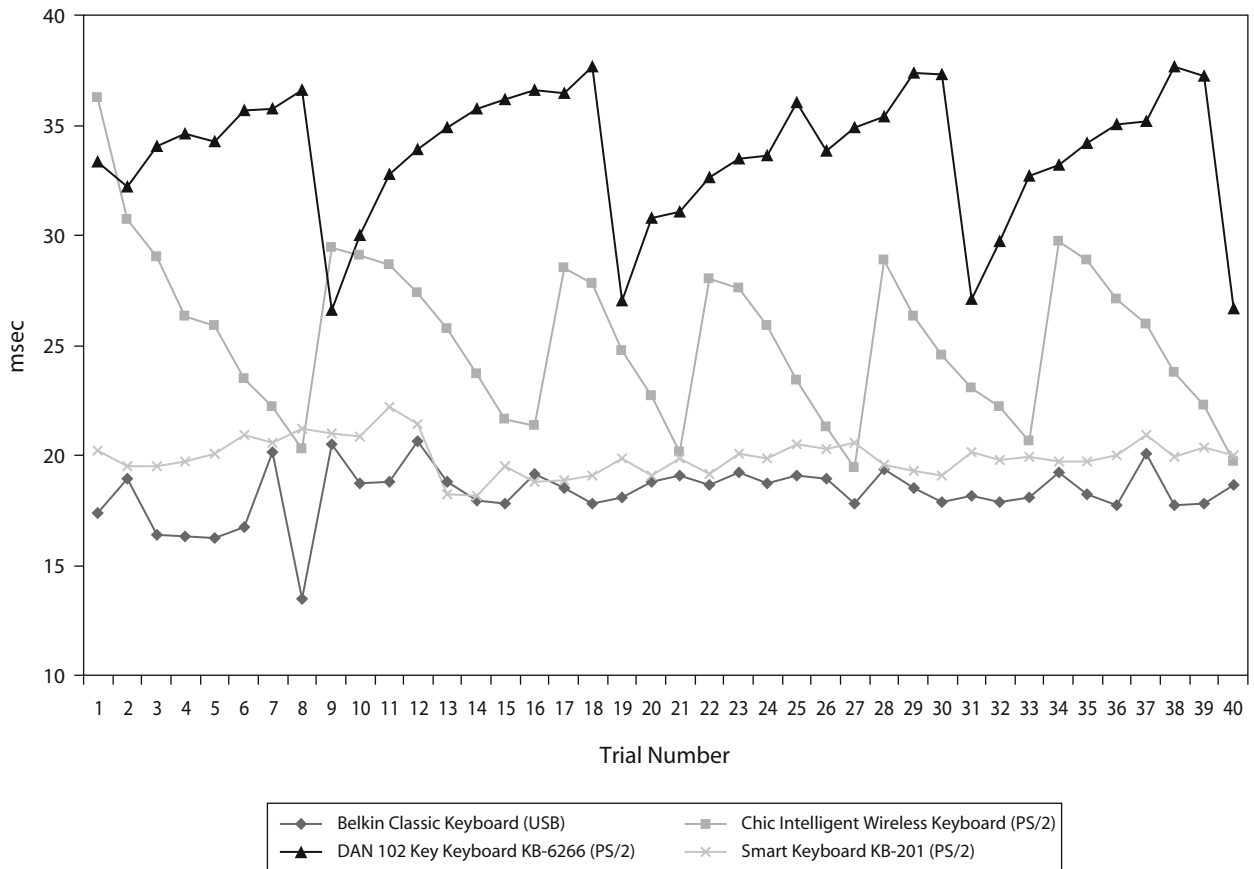


Figure B9. Summary of response time delays for keyboards tested (target = 0 msec).

Table B13
Summary of Response Time Delays
for Keyboards Tested

Model (Interface)	Response Time	
	<i>M</i>	<i>SD</i>
Belkin Classic Keyboard (USB)	18.30	1.29
Chic Intelligent Wireless Keyboard (PS/2)	25.34	3.68
DAN 102 Key Keyboard KB-6266 (PS/2)	33.73	3.08
Smart Keyboard KB-201 (PS/2)	19.94	0.83

Table B14
Computer Speaker Systems
and Interface Mode

Model	Interface
Acer AL718	3.5 mm
Jazz J1321	USB
Philips A1.2FPP005	3.5 mm
Yamaha YST-M20DSP	3.5 mm
Encore P-905U	USB
Labtec LCS1060	3.5 mm
Viglen EZX15F (built into TFT)	3.5 mm
Yamaha YST-M20DSP (DSP active)	3.5 mm

APPENDIX B (Continued)

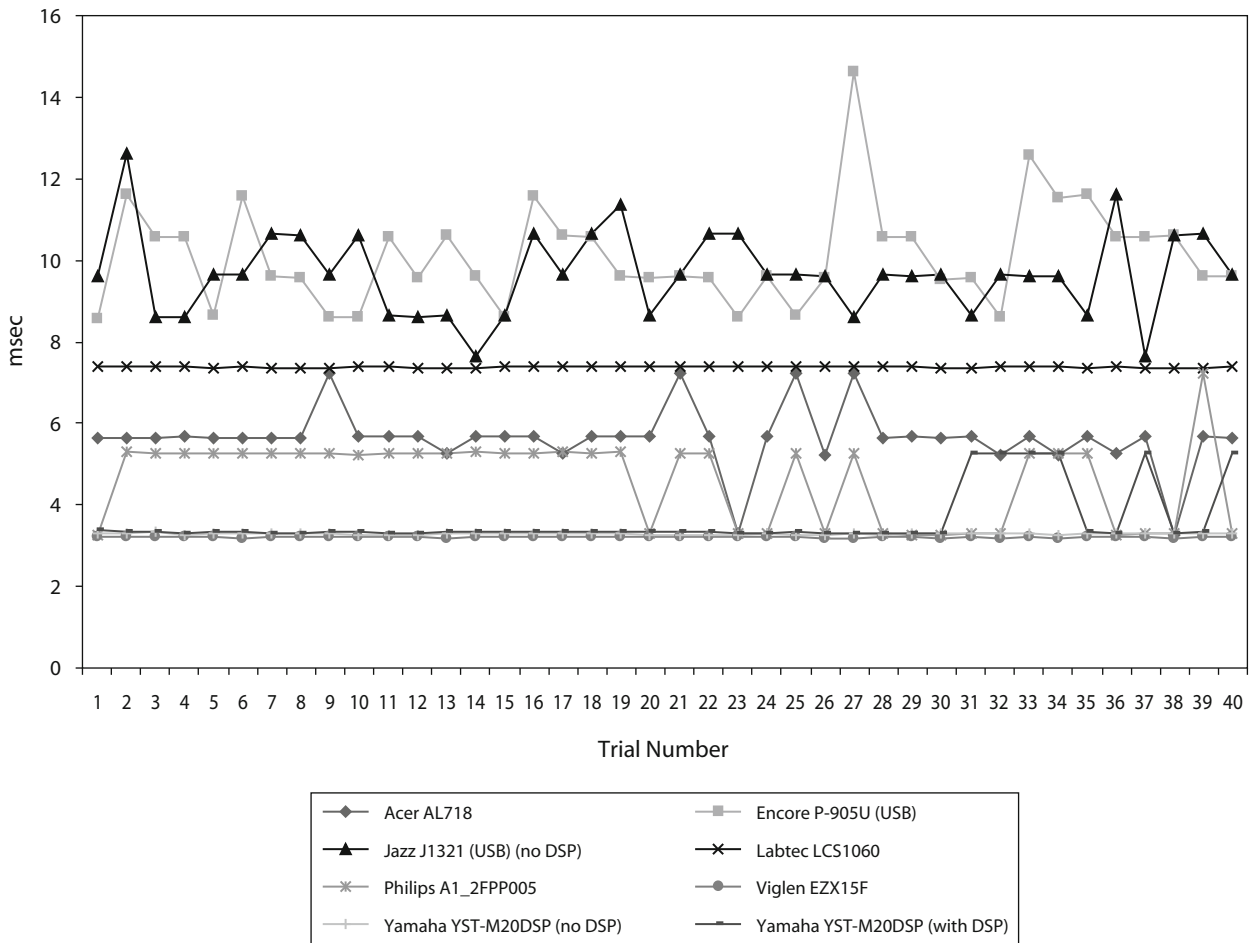


Figure B10. Summary of onset times for speakers tested (target = 0 msec).

Table B15
Summary of Onset Times for Speakers Tested

Model	Onset	
	<i>M</i>	<i>SD</i>
Acer AL718	5.64	0.75
Encore P-905U (USB)	10.12	1.26
Jazz J1321 (USB) (no DSP)	9.68	1.05
Jazz J1321 (USB) (generic DSP)	37.00	1.31
Labtec LCS1060	7.38	0.02
Philips A1_2FPP005	4.62	1.04
Viglen EZX15F	3.21	0.02
Yamaha YST-M20DSP (no DSP)	3.29	0.01
Yamaha YST-M20DSP (with DSP)	3.62	0.70

APPENDIX B (Continued)

Table B16
Summary of Onset Times for Speakers Tested
With and Without DSP Active

Model	Onset	
	<i>M</i>	<i>SD</i>
Yamaha YST-M20DSP (no DSP)	3.29	0.01
Yamaha YST-M20DSP (with DSP)	3.62	0.70
Jazz J1321 (USB) (no DSP)	9.68	1.05
Jazz J1321 (USB) (generic DSP)	37.00	1.31

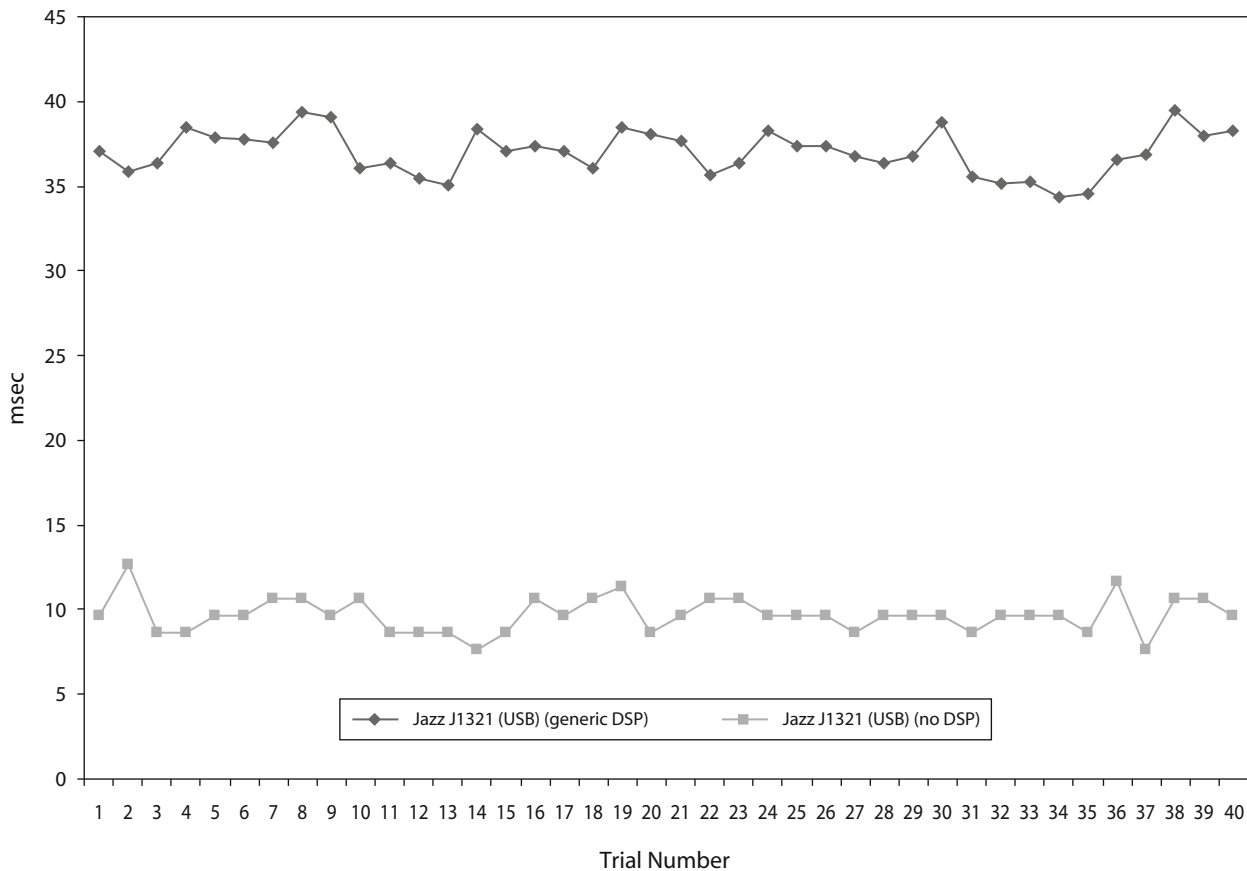


Figure B11. Driver/DSP effects on onset delay (target = 0 msec).

(Manuscript received December 10, 2007;
 revision accepted for publication May 8, 2008.)