11/20/2009

Ms. Kaplan,

The revision of Energy Star programmable home thermostat specifications was discussed widely at the recent Behavior, Energy & Climate Change conference Washington, DC. Harvey Sachs of the ACEEE suggested I should send my comments directly. Please let me know if I'm not following the proper process.

It is encouraging that the draft 2.0 programmable thermostat specifications recognize the importance of usability and suggest evaluation criteria (lines 353-399). However, specifying usability requirements in terms of button presses or other low-level physical and functional properties:

* will not ensure intuitive interfaces
* will entrench current thermostat design features thus discouraging innovation, and
* will render the specification rapidly obsolete

For example, the optimal backlight inactivity time-out (line 379) will vary depending on many interface design choices, and what (if any) user interaction is in progress. Such issues are already addressed by requiring a maximum thermostat power consumption (line 428). If telephone usability requirements were specified in terms of minimum finger openings for rotary dials, mandatory key layouts, and mandatory speed-dial buttons, the Apple iPhone would be deemed non-compliant and to have 'poor usability'.

Instead, I suggest that specifications should follow the performance-based summative usability test method as described in the ISO 9126 standard (http://www.usabilitynet.org/tools/r_international.htm#9126-1). Desired outcomes (effective, persistent consumer adoption and use) should inform specific task performance metrics for live participants to achieve. To ensure fairness and accuracy, usability testing could be performed by an EPA-selected consultant, similarly to Underwriters Laboratories. By focusing on end objectives, the Energy Star programmable specification could include more universal, robust, and device-independent metrics comprising:

1) The thermostat tasks that users must complete (which the Draft specification already includes: setting schedules, modifying schedules, setting energy-saving away mode)
2) The laboratory environment in which the tasks will be performed (e.g. wall mounting height, day/night lighting)
3) The performance metrics to be met for Energy Star acceptance.

Metrics widely used by usability professionals include success rate, error rate, quitting rate, and time taken, as in ISO/IEC 9126-2. Metrics
should be very specific, for example:

"At least 75% of users shall complete programming the schedule and temperature settings in Table A in less than 3 minutes."

"At least 90% of users shall complete setting an energy saving away mode in 30 seconds on their first trial, and in less than 5 seconds on their second trial."
"Trials in low lighting conditions shall take no more than 25% more time than in daytime lighting"

The specification would also specify standard test procedures, such as:

1) What demographic range of users shall be recruited for evaluations (age, education, prior use of programmable thermostats, other relevant demographics)
2) The standard testing protocol (Training time, scripts to be read by test administrators, etc.)

Such performance-based usability metrics are widely used throughout the computing industry. Large thermostat device designers almost certainly already use usability testing in their internal design process. The Federal government has adopted usability evaluation methods and metrics for webpage design
(http://www.usability.gov/methods/test_refine/learnusa/preparation.html
). Notably, hospitals are now beginning to procure medical devices based on usability test outcomes (
http://www.cmaj.ca/cgi/content/full/178/3/273).

The remainder of the technical Draft 2.0 specifications are excellent, particularly requirements for default energy-saving settings. At the same time, I hope you will agree that performance-based metrics are more appropriate to specify usability.

Regards,

-Antony Hilliard, M.ASc.
PhD Candidate, Industrial Engineering, University of Toronto