Audio Engineering Society

Convention Paper

Presented at the 129th Convention
2010 November 4–7 San Francisco, CA, USA

Locating Sounds Around the Screen

David Black[1] and Jörn Loviscach[2]

[1] Hochschule Bremen, University of Applied Sciences, Bremen, 28199 Germany
dblack34@gmail.com

[2] Fachhochschule Bielefeld, University of Applied Sciences, Bielefeld, 33602 Germany
joern.loviscach@fh-bielefeld.de

ABSTRACT

Today's large-size computer screens can display a wealth of information easily enough to overload the user's visual perceptual channel. Looking for a remedy for this effect, we research into providing additional acoustic cues through surround sound speakers mounted around the screen. In this paper, we demonstrate the results of user evaluations of interaction with screen elements using the surround-screen setup. Results of these tests have shown that applying surround-screen sound can enhance response times in a simple task, and that users can localize the approximate origin of a sound when played back with this technique.

1. INTRODUCTION

The continually declining cost per screen pixel in computer displays has meant that both consumers and professionals alike have an increasingly higher amount of screen real estate at their disposal [11], [14]. Whether larger single screens, multiple monitors, or projector displays, the trend is towards more screen space [4]. This has not, however, been accompanied by a wholesale subsequent optimization of this space. Quite to the contrary, in many applications, this explosion in screen space has resulted in the proliferation of more application windows, more screen widgets [14], and — in the case of digital audio workstations (DAW) — more virtual sliders, knobs, and waveform and track displays. Although an ever-increasing amount of information is displayed, the lack of quick and useful access to this screen space leaves much room for improvement.

To help tackle the problem of an abundance of screen real estate with burdensome navigation, we evaluate a system which adapts the surround-sound concept, increasingly applied in both consumer and professional contexts, to aid the user in various modes of interface navigation. Rather than placing sounds around the user in a horizontal circle (as with a standard surround-sound arrangement), we instead place small speakers directly around the bezel of the screen.

2. RELATED WORK

The benefits of such a large display configuration include increased task productivity, higher user satisfaction in completing tasks, and improved recognition memory and peripheral awareness [8]. After using multiple monitors or larger displays, many users adapt to the extra screen real estate and many even claim that they would never return to using a single monitor [9]. However, the large screen area is not without its drawbacks. Roberston et al. [14] identify common problems with large displays, such as losing the mouse cursor and the increase in both the number of visible windows and number of complex tasks. Much progress has been made to help alleviate the problems of large screen spaces and to organize information displayed on this increasingly large canvas. Research ranges from new visual frameworks for desktop organization to spatial auditory displays involving multiple sound sources to 2D-based sound browsing using stansard surround sound.

To improve the targeting of a highly accelerated mouse cursor on a large or wall-size screen, Baudisch et al. [4] present a high-density mouse cursor which helps the user track the position by super-sampling the cursor position and filling in additional cursor images between the actual cursor locations. In contrast to the visual “mouse trail”, this method retains the responsiveness of the cursor and improved performance on a Fitts’ Law targeting task. To help direct user attention on large displays, Khan et al. [11] darken regions outside of the spotlit region of attention, similar to a spotlight used in theatrical productions, which also is used to draw attention to a specific area. New techniques for organizing the information on the large screen space are also gaining traction. Agarawala and Balakrishnan [2] present a redesign of the virtual desktop paradigm with the BumpTop, a prototype that integrates physics-based interaction and visualization techniques that are optimized for pen input. Objects on the desktop may be flung around the screen, stacked, grouped, and filed as they would be on a real desk with accompanying physical objects and their real-life characteristics.

Optimization of the use of large and multiple displays has not been limited to the visual domain alone. In systems for aircraft cockpits and military maritime support, spatialized audio has been used to aid visual search and control tasks. Brock et al. [7] implemented spatialized sound in a watch-station system that comprised three separate displays to reduce the amount of head turns in a dual-task paradigm. Two tasks were displayed on the two displays to the left and right of the center display. By using spatialized audio, the subjects turned their heads significantly fewer times than without any sound to confirm successful data entry. Bolia et al. [6] found that spatial audio was effective in a target-acquisition task for aircraft cockpits involving a range of 180º horizontally and 160º vertically. Their results indicate that both free-field and virtual audio cues significantly decreased search times and did not entail a decrease in the number of correct task responses. Agres et al. [3] have tested head orientation and a virtual ambisonic system and its effects on both simple and complex searching tasks.

Extending the virtual desktop metaphor to the sound domain, Heise et al. [10] present an interface to aid in browsing large sound databases. The zoomable tool uses a flashlight metaphor to scan and quickly audition many sounds on the screen canvas at once. All sounds within the cone of light are played back, and the user may select the size of the flashlight cone to play one or many sounds at once. The sounds are represented by icons, which are grouped according to similarity. Surround-sound modalities are used to help navigate through the large collections of sounds. They have shown that this method results in faster sound file retrieval than standard file-by-file auditioning methods.

3. EXPERIMENTS

The evaluation consisted of a test with eleven subjects between the ages of 25 and 35. Seven were female and four were male. Participants self-reported normal or corrected eyesight and normal hearing. Of the eleven test subjects, two claimed significant experience with audio software. Six test modalities were included, three of which tested the subjects’ reaction time to an auditory and visual notification, and three of which tested the subjects’ ability to localize a sound using the surround-screen technique. The tests were completed using an Apple iMac computer with a 27-inch display. Two pairs of Speedlink Definition SL-8020 powered consumer computer speakers were disassembled, from which the speaker units (6.5-centimeter diameter) were removed from the housings and attached to the corners of the computer display’s bezel with polyethylene adhesive cloth (Fig. 1). The placement of the speakers on the bezel were made to be as close to the screen’s display as possible without obscuring any pixels of the test screen area. A Sinn7 Rox.6 USB interface and the Mac external audio output fed the speakers’ internal amplifiers.

The signal levels of each speaker were equalized before testing. Input was given by using solely a Logitech MX400 mouse. The test software was written in Max/MSP and localization of pixel-to-panning used two equal-power distributions (left-right and top-bottom pairs) [13]. Subjects were each asked to sit with their heads at a uniform distance of 40 cm to the middle of the screen. Although they were given this explicit instruction, subjects did tend to slightly move either forwards or backwards from the original position.

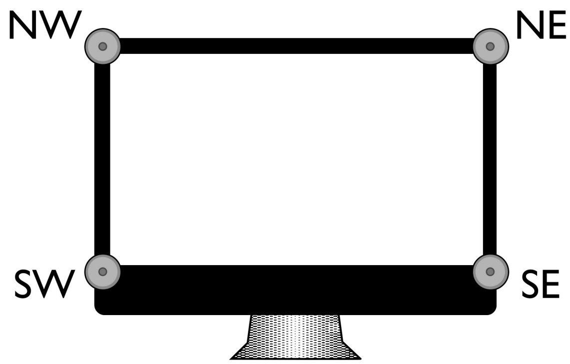

Figure 1. The layout of the 4 speakers, denoted by cardinal directions, affixed to the bezel of the display

3.1. Test Modes

Six modes of testing were completed with each subject. For all modes, the interaction consisted of the playback of a burst of pink noise with a 5 ms attack and 300 ms decay followed by a mouse click by the subject on the display, which had a clickable area of 2516 x 1342 pixels. Subjects completed 20 iterations of each testing mode, resulting in 120 clicks per user. A randomly-timed delay between 3 and 5 seconds separated each of the 20 iterations of the six testing modes. Before each test, subjects were shown a run-through of the complete test to familiarize them with the procedure. To simulate a distracting task, between each testing iteration, subjects were asked to occupy themselves by using the mouse cursor to follow a circle moving in random directions in the middle of the screen. This task forced the subjects to return the mouse cursor to the middle of the screen after each test iteration.

3.1.1. Reaction Time Tests

The purpose of the first three test modes was to determine whether using the surround-screen technique had an effect on the time needed to click on a dialog box. The first three test modes consisted of a 148 x 110 pixel dialog appearing at a random location on the screen along with a white background and the distraction task. The three modes included presenting the dialog box accompanied by no sound, presenting the dialog box accompanied by the noise burst played through all loudspeakers at equal amplitudes (mono sound), and presenting the dialog box with the localized sound. For each iteration of these three testing modes, the distance between the center of the screen (where the user is occupied with the ball-following task) and the position of the mouse as it clicked the dialog box were recorded. The time required to click the dialog box after its presentation was also recorded.

Data collected from clicking times is subject to huge amounts of scatter. It can only rarely be used without smoothing. To smooth the data in a meaningful way, we take Fitts' Law into account: The time taken to move and click can be modeled as a linear function of the information

$$T = \log_2\left(1 + \frac{D}{W}\right),$$

where T is the time taken, D the distance traveled, and W the width of the target. As we are dealing with a two-dimensional target (a button) of 148 x 110 pixels, we set W to the smaller of the two dimensions [12], that is, to 110. Current improvements on Fitts's law suggest more complex approaches [1] to handle this bivariate situation. In the situation at hand, however, the target is almost square and most of the times much smaller than the distance to be traveled.

To cut off outliers and training effects on the start of the learning curve, we only used the second half of the data for evaluation (10 of the 20 clicks for each of the three tests tasks per user). These data are shown in Fig. 2. The data of all participants are pooled to produce enough data points without creating a wearisome test session for each participant.

The trend lines in Fig. 2 are of similar slope but at different heights, indicating that the speed of the motion stays constant but the startup time is lowered by approximately 40 ms when using mono sound as opposed to no sound at all. It is lowered by approximately 130 ms when using surround sound.

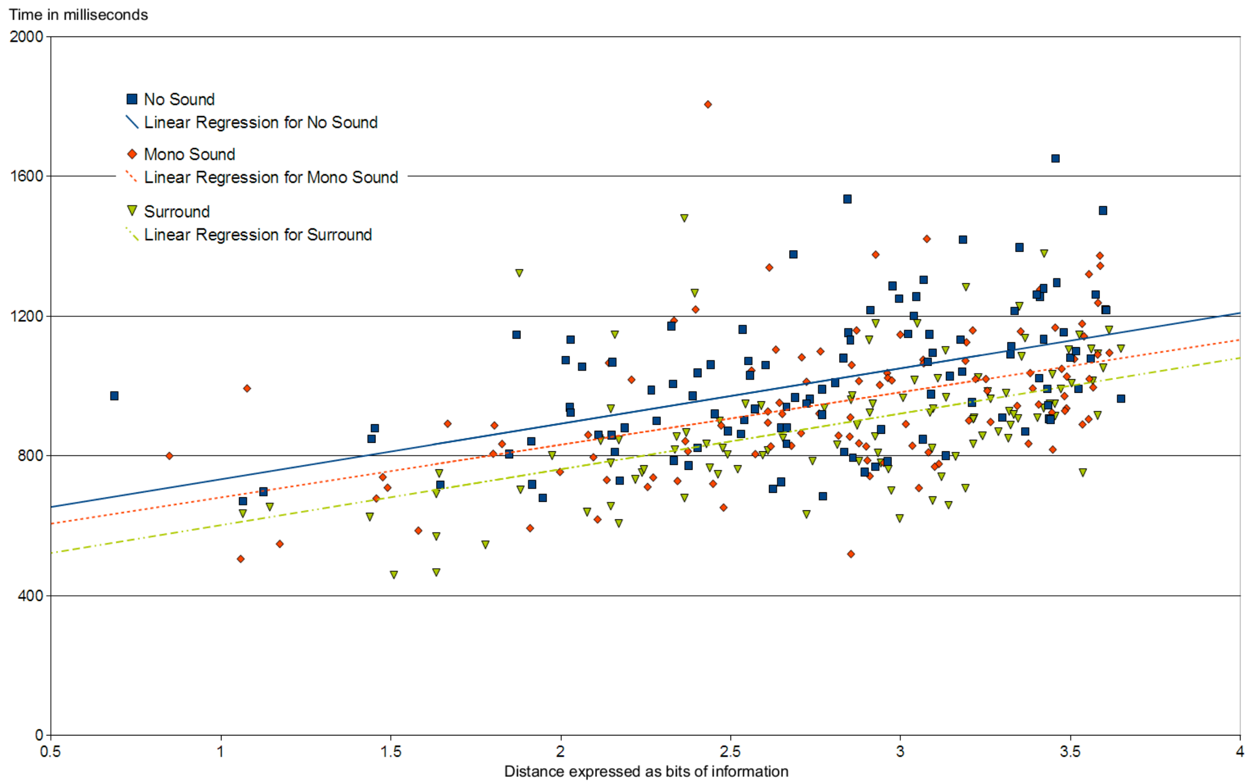

Figure 2: Results of the time-to-click tests, for no sound, mono sound, and surround-screen sound

3.1.2. Localization Tests

The following modes were used to judge the ability of the subjects to place the perceived origin of the noise burst on the screen. The subjects were asked to click on the screen on the point from which the sound was perceived. As in previous modes, between sound bursts, the users were asked to complete the distraction task to that the mouse cursor always began at the middle of the screen. The three test modes included localizing the noise burst solely at the corners of the screen, localizing the noise burst from any one of a series of twelve clickable 110-pixel diameter round buttons in a four-by-three matrix on the screen, and localizing the sound at a random pixel location on the screen (see Fig. 3). For the first two of these modes (corner localization and four-by-three matrix), subjects were presented with four and twelve buttons, respectively. Subjects were asked to click on the button that was perceived to correspond with the localized sound. Clicking on one of these buttons completed the iteration, and the identifier of the clicked button was recorded. For the mode of freely localizing the sound on the screen, subjects simply clicked on the white screen area, after which the location of the click was recorded. For all three of these modes, the time needed to click was not recorded. Each mode was repeated for 20 clicks.

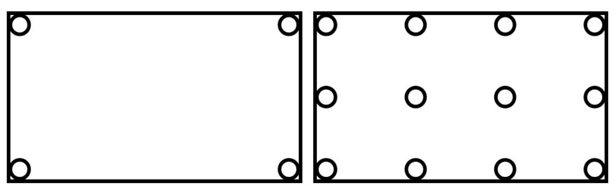

Figure 3: Screen positions of the four and twelve buttons for the first two localization tasks

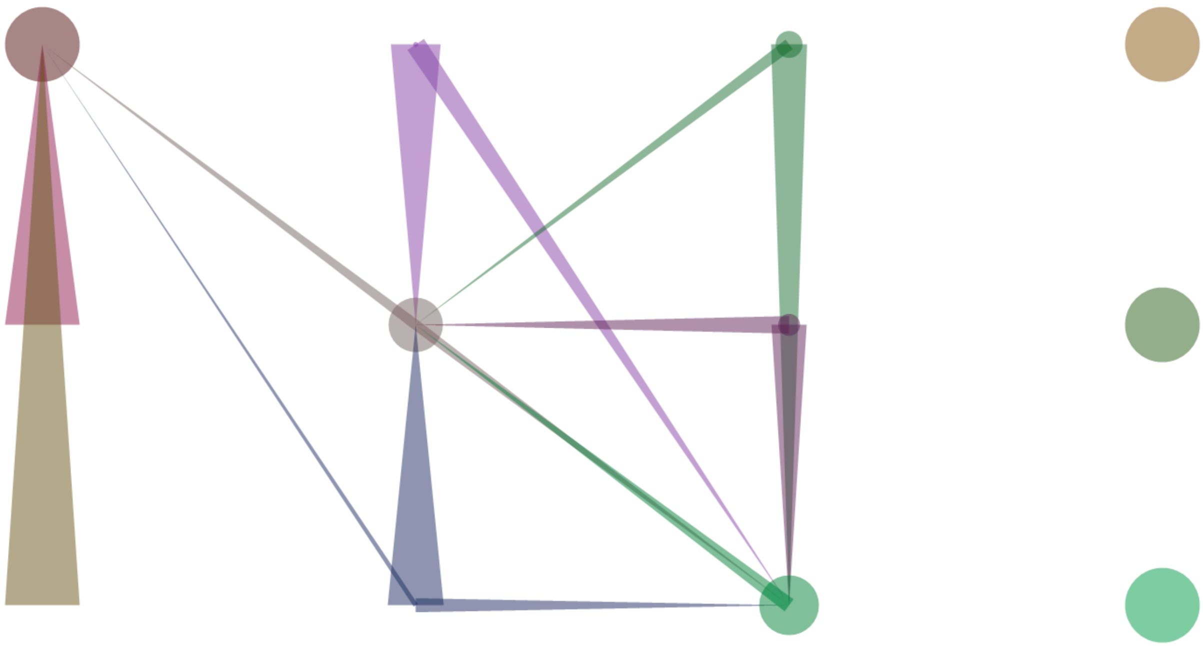

Figure 4: Confusion matrix for the four-by-three button test

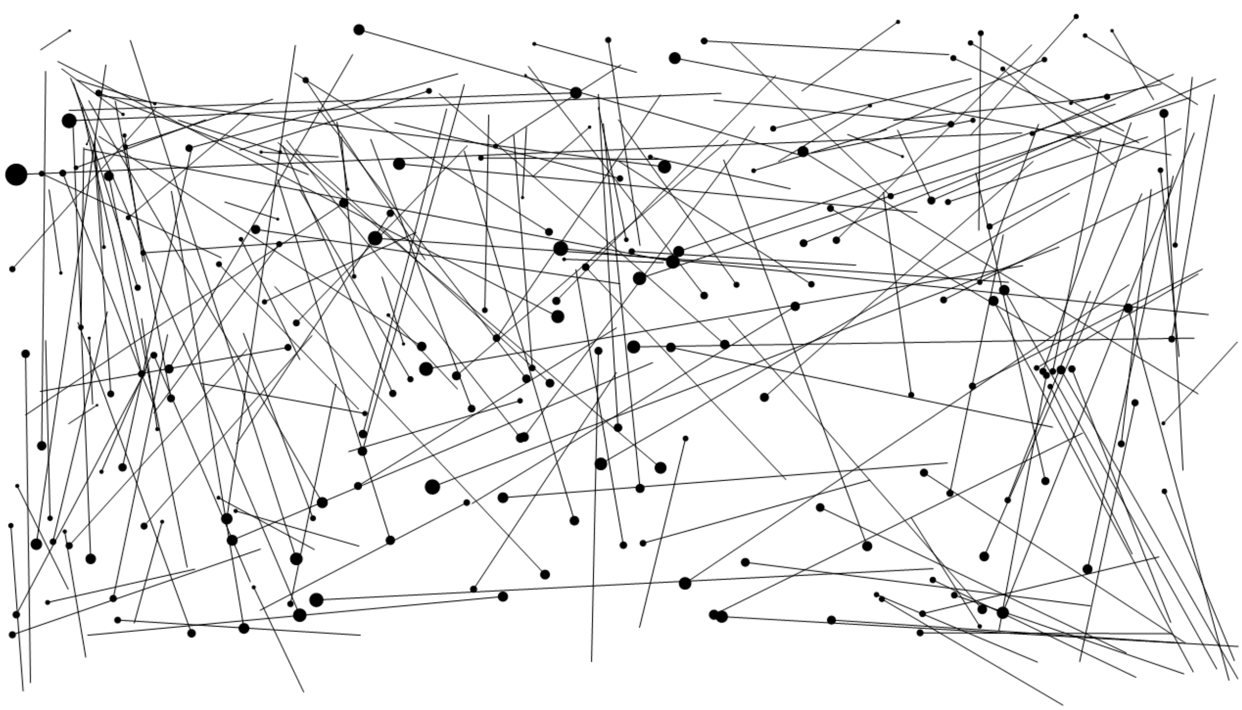

Figure 5: Free-placement test

For the first test involving the buttons placed on four corners of the screen, no tester in all of the 220 test runs confused sounds played back from the left or right as originating from the other direction. However, 10% of the “NE” and “NW” (top of the screen) sounds were mistaken to come from the bottom of the screen. Conversely, 29% of sounds originating from “SE” and “SW” (the bottom of the screen) were mistaken to originate from the top of the screen, showing that there was a clear tendency to place sounds at the top of the screen rather than the bottom.

Fig. 4 represents the results of the second placing test with a graphical confusion matrix. This shows which actual sound position (bases of arrows) is being (mis)interpreted as which screen position (tips of arrows). The widths of the arrows indicate the percentage of cases. The number of perfect hits is indicated by a disk.

As Fig. 4 illustrates, dividing the screen into twelve zones reduces much of the randomness present in the free-clicking test, described below. 48% of the clicks are correct; 42% are off by one position either vertically or horizontally.

For the third test involving freely placing the perceived location of the played sound on the screen, Fig. 5 shows the size, direction, and distribution of the errors of all participants combined. Each dot and tail represents one click. The dot is placed at the intended position and the tail connects this to the actually clicked position. The size of the dot indicates the size of the error.

Apart from obvious outliers, the largest errors occur in the center of the screen. Here, the deviation mostly points upward, supporting the fndings of the first (four buttons on the corners) test.

In all four corners, there is a clear tendency to perceive the sound closer to the corner as is intended. This coherent misinterpretation could be reduced through appropriate predistortion of the position that is fed into the surround-screen system. The average length of the error is 560 pixels, which is 19% of the screen's diagonal and 39% of its height.

4. CONCLUSION

The results of our first three tests show that using surround-speaker clues slightly decreases reaction time to user interface notifications over both modes of monophonic sound and no sound.

The second set of three test results indicate that subjects were able to clearly distinguish left from right, but had a more difficult time in differentiating between top and bottom, favoring placing sounds near the top of the screen. Nonetheless, subjects were able to generally place sounds near their actual source, with accuracy increasing as the number of possible actual sound sources diminished.

Drawing from the results of the tests, surround-screen sound would be a beneficial addition to large or multiple screen display configurations, as it helps to quickly direct the user’s attention to areas of interest on the display, similar to the Spotlight as shown by Khan et al. [11].

In addition to helping users navigate large screen real estates, new applications may benefit from such a technique. Further work is aimed at applying the surround-screen technique to prototype applications which may benefit from such a deployment. Canvas-browsing applications such as that by Heise et al. [10] may benefit from a surround-screen localization more than from a standard surround-sound system, as the sounds on the screen are also placed along the same 2D space. Fragments of audio that are selected may be spatialized to appear to originate from the vicinity of the mouse pointer. Furthermore, standard DAW software could benefit by soloing individual tracks or clips based on where they are placed on the screen. This might aid in tasks of searching for a particular clip of sound in a screen-wide DAW track waveform display.

Future tests should evaluate the effect of different sound generation techniques, varying loudness levels, screen size, design of the test signal, and subject’s distance to the screen. In addition, we plan to test an exaggerated placement of the speakers by moving them further away from the bezel of the screen.

Using cheap, off-the-shelf, consumer-grade speakers, we were able to show that the interaction with the user interface benefits from the surround-screen spatialization technique, and that specific applications may also exploit this concept.

5. REFERENCES

[1] Accot, J., Zhai, S., “Refining Fitts’ law models for bivariate pointing,” in Proc. CHI 2003, pp. 193–200

[2] Agarawala, A. and Balakrishnan, R., “Keepin’ it real: pushing the desktop metaphor with physics, piles, and the pen,” in Proc. CHI 2006, pp. 1283–1292

[3] Agres, K., Topel, S., Spivey, M., Moseson, S., “An auditory twist on visual search using head movement controlled ambisonics” in Proc ICMC 2010, pp. 486–489

[4] Baudisch, P., “Interacting with wall-size screens,” in Workshop on Information Visualization and Interaction Techniques for Collaboration across Multiple Displays (April 2006)

[5] Baudisch, P., Cutrell, E., Robertson, G. “High- density cursor: a visualization technique that helps users keep track of fast-moving mouse cursors,” In Proc. Interact ’03, pp. 236–243

[6] Bolia, R., D'Angelo, W., and McKinley, R., “Aurally-aided visual search in three-dimensional space,” Human Factors, 41(4), pp. 664–669 (1999)

[7] Brock D., Stroup J., and Ballass, J., “Using an auditory display to manage attention in a dual task, multiscreen environment,” in Proc. ICAD 2002

[8] Czerwinski M., Smith, G., Regan, T., Meyers, B., Robertson, G., and Starkweather, G., “Toward characterizing the productivity benefits of very large displays,” in Proc. Interact 2003, pp. 9–16

[9] Grudin, J., “Partitioning digital worlds: focal and peripheral awareness in multiple monitor use,” in Proc. CHI 2002, pp. 458–465.

[10] Heise, S., Hlatky, M., and Loviscach, J., “SoundTorch: quick browsing in large audio collections,” In Proc. 125th Audio Engineering Society Convention, number 7544 (October 2008)

[11] Khan, A., Matejka, J., Fitzmaurice, G., Kurtenbach, G., “Spotlight: directing users’ attention on large displays,” in Proc. CHI 2005, pp. 791–798

[12] MacKenzie, I.S., Buxton, W., “Extending Fitts’ law to two-dimensional tasks,” in Proc CHI 1992, pp. 219–226

[13] MIDI Manufacturer’s association, “Recommended practice, default pan formula”, http://www.midi.org/techspecs/rp36.php (last accessed September 8, 2010)

[14] Robertson G., Czerwinski, M., Baudisch, P., Meyers, B., Robbins, D., Smith, G., Tan, D., “The large-display user experience,” IEEE Computer Graphics and Applications, 25(4), pp. 44–51 (July 2005)