

The Adjustable Grid: A Grid-Based Cursor Control Solution using Speech Recognition

Tarif Haque¹, Emily Liang²

Advisor: Dr. Jeff Gray¹

¹Department of Computer Science, ²Departments of Biology and Anthropology - The University of Alabama

INTRODUCTION

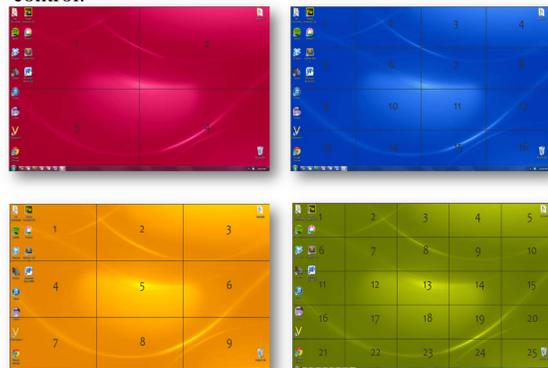
Individuals with motor disabilities often find hands-free, speech-based systems useful because they provide an alternative to traditional mouse-centered navigation. A small number of grid-based cursor control systems using speech recognition have been developed. These systems typically overlay a numbered 3x3 grid on the screen and allow the user to recursively drill the cursor down to a target location by speaking a grid number. Though a 3x3 grid remains the standard, it still remains elusive as to which granularity maximizes performance in specific desktop environments, particularly in regard to time delays and error rates of click tasks. The objective of this research is to develop a grid of adjustable granularity both to compare the efficacy of a variety of grid sizes and to provide users with an alternative to current systems which only offer a single default grid.

1	2	3
4	5	6
7	8	9

The standard overlaid 3x3 grid for speech-based cursor control. To navigate to the red dot, the user says "Three" then "Five" then "Click."

SOFTWARE PROTOTYPE

We implemented a prototype of the Adjustable Grid in Java for Windows 7. Carnegie Mellon's *Sphinx*, an open-source speech recognition toolkit, handled speech recognition functionality. As shown below, the Adjustable Grid generates grid overlays of differing granularities that can be used for speech-based cursor control.



Sample grid overlays generated by the Adjustable Grid. A 2x2 grid (top-left), 3x3 grid (bottom-left), 4x4 grid (top-right), and 5x5 grid (bottom-right) are shown.

FUNCTIONALITY

The Adjustable Grid first asks the user to specify the width dimension and then the height dimension of the grid using speech commands. To illustrate, if the user says "4" then "5" during these prompts, a 4x5 grid is painted on the screen, as shown below.

"one"	"two"	"three"	"four"
"five"	"six"	"seven"	"eight"
"nine"	"ten"	"eleven"	"twelve"
"thirteen"	"fourteen"	"fifteen"	"sixteen"
"seventeen"	"eighteen"	"nineteen"	"twenty"

Dynamic grammar generation using a 4x5 grid. For illustrative purposes, each vocabulary is shown with its respective grid. Generating the vocabulary after the grid has been specified maximizes performance for each individual grid by limiting the number of speech commands that can be spoken and thereby reducing speech recognition errors.

SELECTING A TARGET



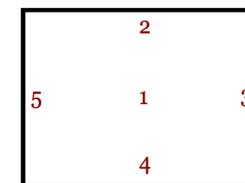
Selecting a target with a 4x3 grid. To click the icon, the user says "Four", then "Four" again, and finally "Click" to select the target.

EXPERIMENTATION

To our knowledge, no previous research compares grids of differing granularities for speech-based cursor control. Two comparison-based studies were designed and performed by the research team. The first study evaluated the performance of pure-resolution grids (e.g. 2x2, 3x3, 4x4) against one another. The second study sought to investigate the potential of mixed resolution grids (e.g. 4x3) by comparing a 4x3 grid to the standard 3x3 grid.

EXPERIMENTAL PROCEDURE

Two members of the research team tested each of the grids on two separate machines. Five targets were placed on the screen, as shown in figure below.



Layout of the experimental screen showing the five locations of the target icons used to test the grids generated by the Adjustable Grid.

TESTING PROTOCOL

First, the user specifies the grid size using speech commands. The selected grid is overlaid on the screen and the user attempted to navigate to the first icon. After clicking the icon, a full screen program launches. The user then navigates to the appropriate location and issues a "click" command to close the program. After closing the first program, the user navigates to the second target, repeating the procedure for icons 2, 3, 4, and 5. For each target, the total time to specify the grid size, open, and close the program is recorded. This five-target procedure is then repeated for all the grid-sizes being tested.

SUMMARIZED EXPERIMENTAL DATA

Table 1. Summarized Pure Grid Performance on Machine I

	2x2	3x3	4x4	5x5	6x6
Time	26.69 (2.12)	25.76 (9.85)	28.65 (5.81)	19.96 (6.14)	18.87 (2.74)
Commands	13.0 (0.00)	10.40 (2.07)	12.60 (2.19)	10.0 (2.92)	8.60 (1.34)
Errors	0.20 (0.45)	2.80 (2.95)	1.40 (0.89)	1.00 (1.73)	0.20 (0.45)

Table 2. Summarized Pure Grid Performance on Machine II

	2x2	3x3	4x4	5x5	6x6
Time	30.00 (1.06)	19.04 (1.23)	28.56 (8.42)	20.79 (2.04)	22.63 (4.07)
Commands	13.40 (0.89)	8.80 (0.45)	10.60 (2.50)	8.80 (0.45)	8.00 (0.71)
Errors	0.20 (0.45)	0.20 (0.45)	1.00 (1.41)	0.00 (0.00)	0.20 (0.45)

Table 3. Mixed-Resolution Grid vs. Pure-Resolution Grid

	4x3 [PC I]	4x3 [PC II]	3x3 [PC I]	3x3 [PC II]
Time	16.44 (1.29)	19.53 (1.2)	25.76 (9.85)	19.04 (1.23)
Commands	9.4 (0.55)	8.4 (0.89)	10.40 (2.07)	8.80 (0.45)
Errors	0.00 (0.00)	0.40 (0.55)	2.80 (2.95)	0.20 (0.45)

RESULTS

PURE-RESOLUTION GRID COMPARISONS

Our results suggest an improvement for 5x5 and 6x6 pure-resolution grids. The 3x3 grid performed well on Machine 2, but not as well on Machine 1. The 5x5 and 6x6 grids, however, consistently performed well across both machines. Even with their increased vocabulary size, the 5x5 and 6x6 grids appear to work as well, if not better, than the 3x3 grid.

Across the experiments, the relatively high standard deviations for selection time suggest existing speech recognition tools continue to be unreliable. The 2x2 grid, with a small vocabulary of four words, appeared to be the only grid that consistently performed at error-free levels.

MIXED-RESOLUTION GRID COMPARISONS

Our findings show clear potential for mixed-resolution grids. The 4x3 grid performed better, if not as well, as the 3x3 grid on multiple machines. On Machine I, the 4x3 grid performed significantly better than the 3x3 grid, with a 36% improvement in completion time at error-free levels. In addition, the low standard deviations of the 4x3 grid measurements compared to the 3x3 grid shed light on the grid's reliability.

CONCLUSION

We suggest that future grid-based cursor control systems focus less on a singular 3x3 grid, and take a dynamic, adjustable approach that offers differing granularities, an implementation that grants the user increased flexibility. Multiple grids have shown to match and surpass the performance of the 3x3 grid. As user interfaces become more expansive and demanding, relying on a single grid will be limiting to users.

ACKNOWLEDGEMENTS

This research was made possible by The University of Alabama and The National Science Foundation.

