September 2021

Multi-modal Gesture Triggering for Touchscreen Input

Thomas Buckley

Robert Mickle

Abraham Murray

Tayeb Al Karim

Hiu Ying Chan

See next page for additional authors

Inventor(s)
Thomas Buckley, Robert Mickle, Abraham Murray, Tayeb Al Karim, Hiu Ying Chan, and Maria Cirimele

Multi-modal Gesture Triggering for Touchscreen Input

ABSTRACT

Stylus apps support handwritten gestures for common actions by providing a menu to distinguish strokes from gestures. However, switching between the menu and the canvas can create context-switching friction. Moreover, attempting to reduce friction by eliminating the context-switch can lead to gestures being triggered unintentionally, unreliably, or unpredictably. This disclosure describes a multimodal gesture interface with a touchscreen that, with user permission, interprets spoken input provided by the user to confirm a gesture and its intended behavior. For example, after drawing a circle the user can say “select” to trigger the select gesture for strokes that the circle bounds. The user can alternatively say nothing to retain the circle as a stroke. As another example, the user can speak a different command, e.g., “delete” to erase strokes that the circle intersects.

KEYWORDS

- Stylus input
- Touchscreen
- Spoken command
- Gesture input
- Gesture disambiguation
- Sketching app
- Multimodal interface
- Speech command
- Natural language processing

BACKGROUND

Applications that make use of a stylus or allow finger-based input, e.g., a sketching app, a note-taking app that accepts handwritten input, etc., recognize multiple types of input such as stroke, gesture, etc. performed using the stylus. Such input can also be provided via a mouse or drawing tablet. A stroke is a trajectory traced by the stylus (or finger) on a touchscreen. A stroke can leave behind an electronic mark, such as a squiggle, on the touchscreen. A gesture is an action, e.g., select, move, erase, rotate, rescale, etc., on a stroke.

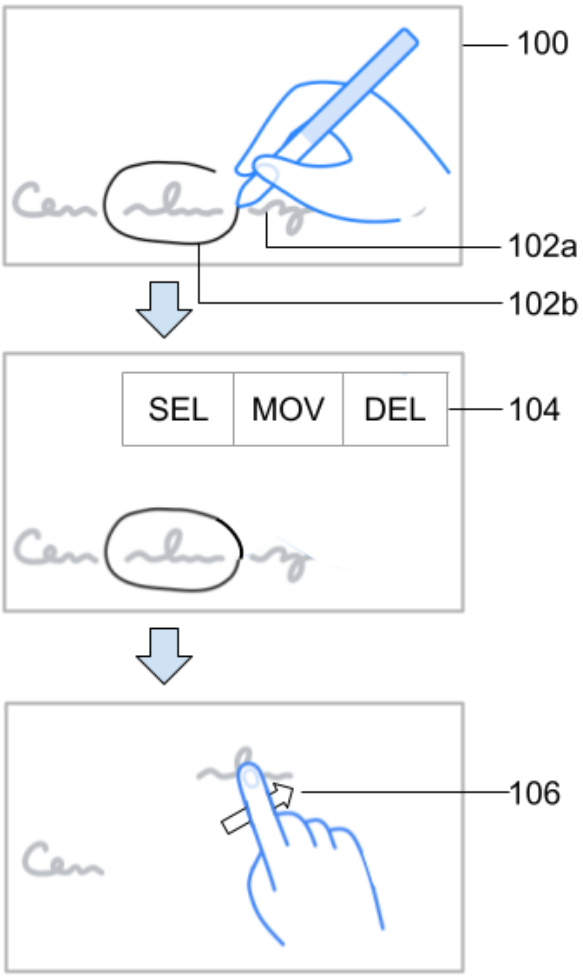

Fig. 1: Stroke and gesture

Fig. 1 illustrates an example sequence of input operations performed by the user on a touchscreen. The user enters a series of handwritten strokes (102a) on a canvas (100) to compose a handwritten note. To move a section of the note, the user marks the section by circling it (102b). At this point, it is unclear if the circle is intended as a stroke or a gesture. Note-taking or drawing apps typically pop up a menu (104) that offers gesture options, e.g., select, move, delete, etc. Sometimes the menu is not a popup menu but provided as a continuous presence at an edge of the canvas. If the circle is intended as a gesture (as it is, in this example), the user selects the appropriate option (MOV) from the menu to move the selection (106). Upon completion of the operation, the circle disappears.

Such switching of context between the canvas and the menu to disambiguate stroke from gesture is strenuous, especially when performed multiple times, as is likely the case for any note or drawing of reasonable complexity. While some stylus apps attempt to support hand-drawn gestures for common actions by providing edit modes that distinguish strokes from gestures, the switching of modes can create friction. Moreover, attempting to reduce friction by eliminating the mode switch can lead to gestures being triggered when not intended, or unexpected application behaviors triggered by the gesture. Gestures sufficiently unusual to be easily disambiguated from strokes are difficult to learn and awkward to trigger. Stylus apps that use heuristics to disambiguate stroke from gesture, e.g., by using timing information, can be unreliable and act unpredictably.

DESCRIPTION

This disclosure describes a multi-modal gesture interface that uses (with user permission) voice to confirm a detected gesture and corresponding intent. For example, after drawing a circle the user can say “select” to trigger the select gesture for strokes that the circle bounds. The user

could alternatively say nothing to indicate that the circle is intended to be a stroke (perhaps part of a diagram). As another example, the user can speak a different command, e.g., “delete,” to erase strokes that the circle intersects.

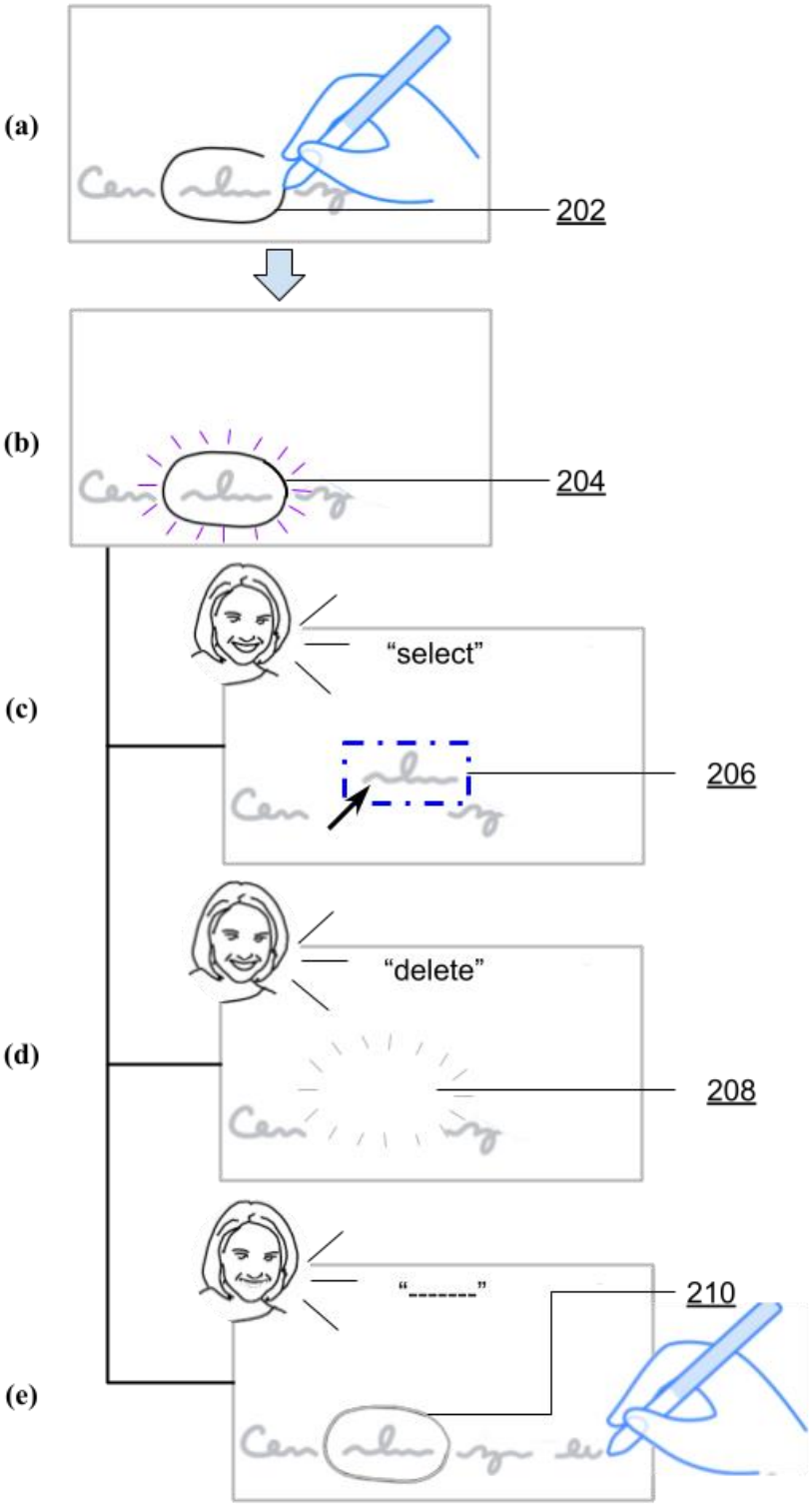

Fig. 2: Multi-modal gesture triggering for touchscreen input

Fig. 2 illustrates an example of voice input for gesture interpretation. The user draws a trace (202, Fig. 2a) in the form of a circle. The trace is initialized to a state intermediate between stroke and gesture (204, Fig. 2b). The intermediate state of the trace is herein pictorially depicted by purple rays; an actual app can use any suitable visual indication of intermediate state.

In the scenario illustrated in Fig. 2(c), the user speaks the word "select," causing the strokes within the circular trace to get selected (206). In the scenario illustrated in Fig. 2(d), the user speaks the word "delete," causing the strokes within the circular trace to get erased (208). In the scenario illustrated in Fig. 2(e), the user says nothing and continues to provide stylus input, causing the circular trace to be recognized as a stroke (210) that is a part of the handwritten text.

Additional examples of spoken commands interpretable as gestures include "duplicate," "rotate," "move up-left," "thicken stroke," and indeed, any menu item. A gesture can have a default value, e.g., "select," until modified by user command. Techniques from natural language processing can be used to interpret the user commands, for example, the phrases "erase," "delete," "remove the stuff inside the circle," etc. can all be interpreted as the same command. In this manner, rather than have the user learn pre-defined gestures to achieve particular stroke manipulations, the user can simply provide natural language spoken input to manipulate strokes on a touchscreen.

Aside from speech, other modes of interaction with the touchscreen can include touch-by-finger; (with user permission) visual gestures detected by a camera; etc. For example, long-press by finger can indicate selection; rapid left-right movement of a finger (as detected by a camera) can indicate deletion; etc. In general, the described techniques open up multiple modes of interaction with a device, including stylus, touch, speech, visual, etc.

Further to the descriptions above, a user is provided with controls allowing the user to make an election as to both if and when systems, programs, or features described herein may enable the collection of user information (e.g., information about a user's touch input, spoken input, commands, gestures, or a user's preferences), and if the user is sent content or communications from a server. In addition, data is treated in one or more ways before it is stored or used so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user. Thus, the user has control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes a multimodal gesture interface with a touchscreen that, with user permission, interprets spoken input provided by the user to confirm a gesture and its intended behavior. For example, after drawing a circle the user can say "select" to trigger the select gesture for strokes that the circle bounds. The user can alternatively say nothing to retain the circle as a stroke. As another example, the user can speak a different command, e.g., "delete" to erase strokes that the circle intersects.