# Explorations and Experiments on the Use of Auditory and Visual Feedback in Pen-Gesture Interfaces

**DIKU Technical Report 04/16**[1]

December 14th, 2004

**Tue Haste Andersen**
Deptartment of Computer Science
University of Copenhagen
Universitetsparken 1, DK-2100 Copenhagen, Denmark
haste@diku.dk

**Shumin Zhai**
IBM Almaden Research Center
650 Harry Rd. NWE-B2
San Jose, CA92120, USA
zhai@us.ibm.com

**ABSTRACT**
We explore the use of auditory feedback in pen gesture interfaces in a series of informal and formal experiments. Initial iterative studies showed that gaining performance or learning advantage with auditory feedback beyond the obvious is difficult. To establish a baseline, Experiment 1 formally evaluated gesture production accuracy as a function of auditory and visual feedback. Size of gestures and the aperture of the closed gesture were influenced by the visual or auditory feedback, while most other features were not. Design implications from this experiment were taken in developing a revised form of EdgeWrite that was proved successful. Experiment 2 focused on the emotional and aesthetic aspects of auditory feedback in pen-gesture interfaces. Participants' views with regard to the dimension of being wonderful and stimulating was significantly higher with musical feedback. Our exploration points to several general rules of auditory feedback that may serve as a foundation for future research and development.

**Author Keywords**
Audio, auditory interface, sound, music, gesture, pen, text input, feedback.

**ACM Classification Keywords**
H.5.2. Information Interfaces and Presentation: User interfaces.

**INTRODUCTION**
This paper deals with auditory displays and pen-gesture interfaces. Due to the increasing importance of mobile forms of computing that lack mouse and keyboard input capabilities or large visual screens, research on these two topics is needed more than ever.

Developing auditory interfaces has long been a captivating research topic in human-computer interaction [4]. It is both an intriguing and difficult research topic. Successful examples of sonification, such as monitoring patients breath or heart beat by auditory interfaces [17], do exist. Auditory feedback in common user interfaces, however, rarely goes beyond the simplest and most obvious forms, despite decades of research.

The current study focuses on coupling sound patterns with pen-gestures such as Graffiti, Unistroke [13], marking menus [15] and SHARK [29]. Several observations motivated us to focus on sound and gestures. First, pen-gesture interfaces are increasingly popular due to mobile and other non-desktop computing forms. Second, it is highly desirable to minimize the visual demand of pen gestures in a mobile computing environment [13]. Third, auditory modality is particularly sensitive to the rhythm and patterns [6] that pen gestures often involve. For example, although it is difficult to tell which tone corresponds to which digit in modem dialing, one could more easily tell if the entire sequence is correct based on the overall rhythm. Another example is the game "Simon," in which the sound pattern may help the player memorize the previous sequence of keys. The powerful coupling between sound and kinesthetic patterns is most evident in musical activities such as dancing and

[1]This report has been submitted for publication outside of DIKU and will probably be copyrighted if accepted for publication. It has been issued as a Technical Report for early dissemination of its contents. In view of the transfer of copyright to the outside publisher, its distribution outside of DIKU prior to publication should be limited to peer communication and specific requests. After outside publication, requests should be filled only by legally obtained copies of the article (e.g. payment or royalties).

instrument playing. It is difficult to imagine dancing without sound patterns (music).

Our research questions are whether and how sound feedback can be created to make pen gestures: 1. more memorable or learnable, 2. more accurately produced for reliable recognition, and 3. more fun, pleasant or engaging to use.

## RELATED WORK AND INFORMAL EXPLORATION

Understanding these research issues requires exploring a vast design space. We started this research with a combination of literature analysis and iterative cycles of design and informal testing.

Early guidelines on audio signal design can be found in Deatherage [7]. In general, absolute position and spatial information are more difficult to represent in audio than in visual display [26]. In contrast, time critical information and short messages can often be more effectively conveyed in audio than in visual displays. In the more recent HCI literature, a number of different paradigms for adding sound to interfaces have been proposed, most notably Auditory Icons [11] and Earcons [2]. Both of these approaches have been primarily concerned with providing feedback about the *state* of a system or the *product* of an interaction transaction.

Providing audio feedback on the product of gestural interaction is relatively easy. Using the SHARK shorthand-on-stylus keyboard text input system [29,14] as an example, we coupled the word SHARK recognized to a speech synthesizer, making the system pronounce each recognized word. With such a function the user did not have to look at the screen to confirm whether the system has recognized the user's gesture as the intended word. Another type of simple and useful state feedback is auditory warnings when the pen deviates from the expected writing area.

To couple sound feedback to the *process* of gesture articulation is both more intriguing and more challenging. One possible gain of coupling sound with gesture movement is to help the user make corrections to the pen's trajectory during articulation or to help users remember the gesture better. The sound patterns therefore have to be highly *discriminable.* We first coupled the articulation of the SHARK shorthand (i.e. the pen movement in the writing area) with a continuous complex tone. The amplitude and frequency of the tone were mapped to the x and y position of the pen tip respectively. This produced discriminable tones for different gesture patterns. However, as in many research audio interfaces, the sound produced by this mapping was not *pleasant*, and it was difficult to imagine prolonged use with this type of feedback. A few design iterations led us to a different approach that mapped the direction of the pen movement to a discrete space of sampled musical instrument sound. Moving the pen in each direction would result in a unique instrument sound. A gesture of a particular movement pattern, such as a triangle shape, sound distinctly different from another, such as a squared shape. This scheme were both enjoyable and discriminable (Clip[2] 1).

To test whether this audio feedback method could help in learning new pen gestures, we tested it with a blind participant using a digital tablet as the pen sensor. The participant was a computer scientist familiar with user interface technologies. A dozen pen gesture patterns were demonstrated to the participant by either holding his hand or letting him trace trajectories carved on a piece of cardboard overlaid on the digital tablet. The participant then practiced reproducing these gestures. It was evident that he had difficulty with memorizing these gestures in a short period of time. Contrary to the common myth that a blind person has more developed hearing capabilities, he tended to memorize the gestures' spatial rather than sound patterns. Considering that we first showed the gestures to him by spatial information (e.g. tracing carved templates), we followed that study with sighted users in which more emphasis was placed on the audio channel in the initial introduction of each gesture. We designed a number of basic pen gesture patterns consisting of straight lines of four directions only (up, down, left, and right); each movement direction was associated with one musical instrument. Participants were acquainted with these mappings first. Each gesture pattern was then presented to the participants by playing the corresponding sound pattern, rather than by visual appearance. At first it was hard to learn the gestures from the sound; it took many trials for the participants to figure out the gesture pattern from the sound pattern played. When the pattern was produced correctly, the participants tended to again remember the pattern by its spatial shape (e.g. a triangle) rather than its corresponding sound. These two informal experiments suggest that the improvement in memorability of gestures from sound feedback is unlikely to be immediate or dramatic. It is possible that people with musical experience may gain more advantage in this regard, but it often takes years of practice to master an acoustic instrument.

Another possible benefit of coupling sound feedback to gesture lies in the gesture production process. Analysis of the literature suggests that this is both possible and difficult. One positive evidence in this regard came from Ghez and colleagues [12] who, in an experiment, added continuous sound feedback to the limb movements of participants who lack proprioception. The sound used was continuous complex tones with varying amplitude and pitch. It was found that with joint rotation and timing information encoded as auditory feedback, the participant was able to perform a certain arm movement with the same timing and precision as when having visual feedback only. This research suggests that providing auditory feedback at the process level may help to reinforce movements used to produce certain gestural patterns.

[2]Although the paper is self-contained, “Clip” numbers are provided as pointers to separate A/V supporting materials.

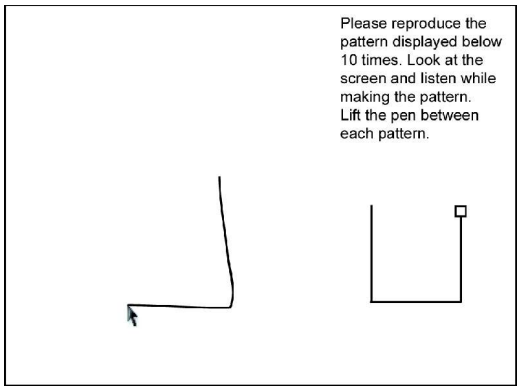


**Figure 1. Screenshot (illustration) of the software used in baseline experiment.**

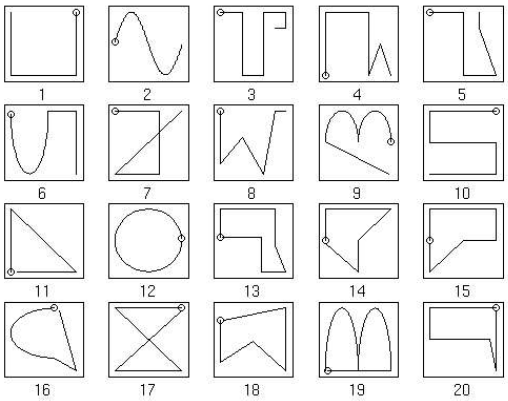


**Figure 2. Gestures used in experiment. Note that gesture 11-20 start and end at the same point.**

On the other hand, there are also reasons to believe auditory (or visual) feedback might be too slow to assist gesture production [24]. The handwriting literature also suggests that gesture production has a strong central representation that is not influenced much by peripheral feedback [28, 16].

The conflicting evidence in the literature about the possible role of auditory feedback requires a more formal and quantitative experiment on auditory feedback of pen-gesture production.

## EXPERIMENT 1: VISUAL AND AUDITORY FEEDBACK IN PEN GESTURE PRODUCTION

This experiment aims at providing baseline measurements on the impact of auditory and visual feedback on gesture *reproduction* accuracy. It will be informative to future work in this area to have a comparative examination of auditory feedback relative to its visual counter part. Another purpose of this study was to identify the roles visual feedback may play in pen gesture and if they could be replaced or avoided altogether.

### Task

The experimental task was to reproduce a set of gesture patterns as accurately as possible at a speed comparable to writing a character with Graffiti. For each pattern, the participant was first shown its shape onscreen, as illustrated in Figure 1. The participant then practiced drawing the pattern with both visual and auditory feedback. After practicing the gesture enough times to be confident in reproducing the gesture without looking at the template, the participant was asked to reproduce the gesture 10 times in various feedback conditions. For the conditions without visual feedback the participants were instructed to close their eyes so that they could not receive visual feedback by looking at their hand. In the conditions where no auditory feedback was given, a noise signal was played to the participant's headphone to mask the sound produced when moving the pen over the surface of the digitizer tablet.

A set of 20 gestures of varying complexity, as shown in Figure 2, was used. Half of the gestures start and end in the same position – we call these closed gestures. The other half were open gestures.

A total of nine people, five men and four women, from 25 to 43 years of age, participated in the experiment.

### Conditions

The reproduction of the gestures was performed in the presence or absence of visual and auditory feedback, resulting in four (2x2) conditions: none, visual, audio, and visual + audio. The order of the four conditions was balanced across participants.

Providing visual feedback was straightforward: the pen trace was displayed on the screen. The user could see the pen movement as well as its entire history instantaneously.

As discussed, the design of auditory feedback for gesture is much more complex. Since the purpose of this experiment was to measure the strongest *possible* impact on gesture reproduction accuracy, we made the auditory feedback as easily discriminable and as instantaneous as possible, without considering the aesthetic aspects. Based on our experience in the exploratory phase of this research, we chose a continuous sound feedback method which mapped the pen's position relative to its starting point to a number of perceptually sensitive parameters. A complex tone was used as the basis of the sound synthesis. Amplitude was mapped to the speed of the pen, so that a fast pen speed created a loud tone. The mapping added rhythmic structure to the sound, and made the sound less intrusive when the pen was at a stop. Vertical position of the pen was mapped to the fundamental frequency of the tone, the most important parameter to the perception of a complex tone [20]. The horizontal position was mapped to the number of overtones: creating more pure tones on the left and richer sound on the

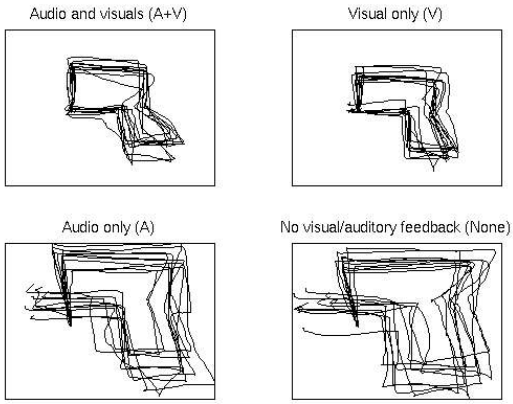


**Figure 3. Example of gesture performed in the four different feedback conditions by a *participant* (cf template 13 in Figure 2).**

right. *Jitter,* i.e random frequency variations of each partial, and inharmonicity [10] were used to further increase perceptual difference in the horizontal dimension. Another perceptual parameter, brightness, was not used since it is known to be confused with pitch [22]. The complex tone can be described as the sound pressure level $s$ at the discrete time $t$ by the following function:

$$s(t) = \sum_{n=1}^{M} 1/n \sin(U F_0 i n t) \quad (1)$$

where $M$ is the number of partials, $U$ is an evenly distributed random number between 0.9 and 1.1, representing the jitter, $F_0$ the fundamental frequency, and $i$ is inharmonisity.

Another perceptual parameter used to encode the horizontal position of the pen was the stereo sound source. If the pen was moved to the left of the starting point, the complex tone was perceived as coming from the left, whereas when it was moved to the right, the tone was perceived as coming from the right. The perceptual direction of the sound source was created by filtering the resulting complex tone with a Head Related Transfer Function obtained from [1], using a Finite Impulse Response filter.

The resulting auditory feedback used in this experiment was thus designed to make different spatial patterns highly discriminable in the auditory space (Clip 2).

### Results and Analysis

The dependent variables we extracted from the pen traces in the experiment included the following:

1. Size of the gesture
2. Aperture: the distance between start and end position for closed gestures

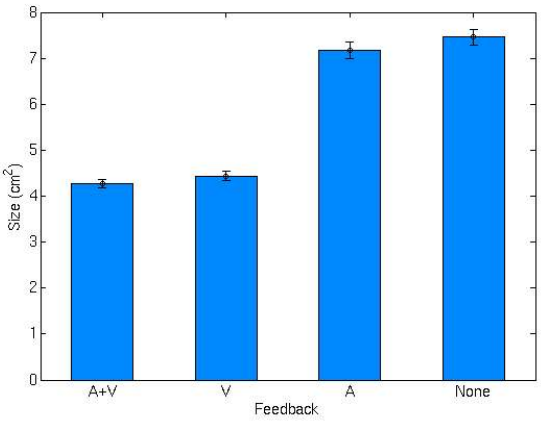


**Figure 4. Size of gestures in four conditions.**

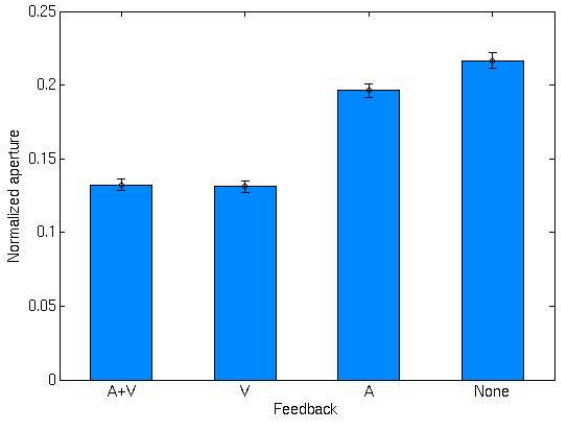


**Figure 5. Aperture between start and end position of the closed gestures.**

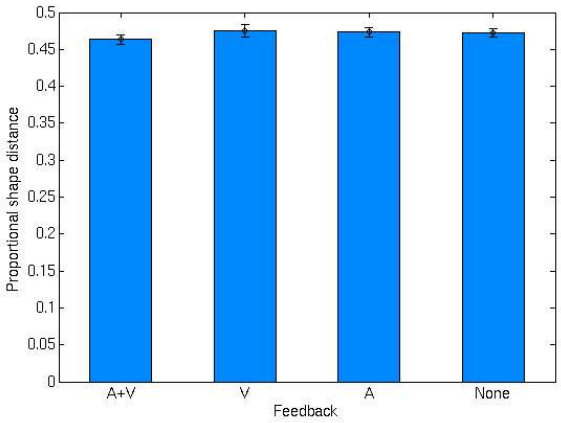


**Figure 6. Proportional shape distance.**

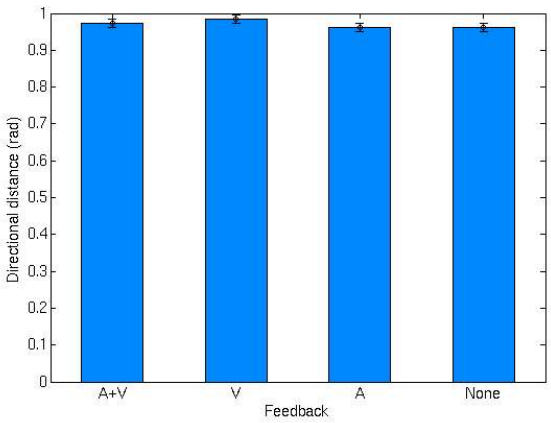


**Figure 7. Directional distance.**

3. Shape distance between the articulated gesture and the template pattern
4. Directional difference between the articulated gesture and the template pattern, for gestures composed of straight lines only

Computationally, the template pattern was represented as a set of consecutive points, $p(n)$, $n=1...M$, where $M$ was the number of sampled points in the pattern. When computing measure 2 to 4, the size of the articulated pattern and the template pattern were first normalized to the same scale and their start positions were aligned.

Within-subject repeated measurement ANOVA analysis was conducted on the four dependent variables, with the feedback conditions as the independent variable with four levels: V+A, V, A, None. We further performed Least Square Distance (LSD) post hoc pairwise comparisons between these levels. Alternatively, we also treated audio feedback and visual feedback as two independent variables each with two levels (presence/absence) and evaluated their main effects and interaction. In general there was no difference in conclusions in the two analysis approaches, and we only report the former, unless the latter was conceptually more informative.

An example of the gestures made by a participant in the four feedback conditions is shown in Figure 3. The example illustrates some of the findings in the statistical analysis described in the following subsections.

*Size*

The size of a gesture is defined as the area in $cm^2$ of the bounding box of the gesture. This variable changed with the feedback conditions significantly (Figure 4): $F_{3,24}=12.5$, $p<0.001$. Post hoc tests showed significant differences between the visual conditions (V, V+A) and the non-visual conditions (A, None) at $p<=0.01$ level. In handwriting literature the effect of feedback on writing size has been observed in the visual channel by van Doorn and Keuss [8], who demonstrated that handwriting without visual feedback was larger than with visual feedback. In our experiment, pairwise comparison of the audio condition (A) with the no-feedback condition (None) was near significance (p=0.076), although the magnitude was small (Figure 4). The difference between A+V and V was not significant.

*The aperture of closed gestures*

For the 10 closed patterns, the Euclidean distance between the start and the end position of the articulated gestures indicated the ability to return to the same point under various feedback conditions. Returning to a past point requires accurate reproduction of the order and direction of each segment as well as the relative proportions of the individual segments in a gesture. It reflects the same type of ability to cross the t's and dot the i's in handwriting research [23].

Figure 5 shows the aperture results in the four conditions. A statistically significant difference was found: $F_{3,24}=14.08$, $p<0.001$. Post hoc tests revealed a significant differences between all pairs of conditions at $p<=0.015$ level, except between V and V+A conditions. Both visual feedback and audio feedback helped to reduce the gap between the start and the end positions, although the impact of audio was much smaller and only significant when no visual feedback was present. One plausible cause was that the audio condition (A) also helped reducing the aperture of the closed gesture in that we used stereo directional sound. One would hear the change of sound direction as soon as the the pen tip passed its starting position in the lateral dimension.

*Proportional shape distance*

To reflect the overall match between the gestures and the templates, we measured "proportional shape distance" $d$, defined as the mean Euclidean distance between a point in the articulated gesture and a corresponding point in the template pattern. If

the template pattern is sampled to points $q\ (n\ ), n\ =0\ldots K-1$ equally spaced along the length of the pattern, and the articulated gesture is sampled at the same number of points ($g$), at the same interval, the proportional shape distance $d$ is defined as:

$$d=\frac{1}{K}\Sigma_{n=0}^{K-1}\sqrt{(q(n)_x-g(n)_x)^2+(q(n)_y-g(n)_y)^2} \quad (2)$$

where $q(n)_x$ and $q(n)_y$ denotes the $x$ and $y$ coordinate of the point $q(n)$ respectively, and $g(n)_x$ and $g(n)_y$ denotes the $x$ and $y$ coordinate of the point $g(n)$ respectively. This measure has been used as the most basic gesture recognition metric in systems such as SHARK [29].

In the experiment, no significant difference was observed between the feedback conditions by the shape distance measure: $F_{3,24}$=0.626, $p$=0.605. (Figure 6).

*Directional difference*

Directional difference describes how well the directional movements in the articulated gesture match the directions of the lines in the template pattern, $\measuredangle\vec{p}\ (n\ ), n\ =0\ldots M_{\ p}$ . The measure preserves the order of the directional movements when comparing the directions of the two patterns, but the extent (length) of each movement is discarded. To achieve this, the directional difference is calculated based on segmentation of the articulated gesture into a list of directions, $\measuredangle\vec{g}(n\ ), n\ =0\ldots M_{\ g}$ . Because the two lists of directions are not necessarily of equal length $(M_{\ p}\neq M_{\ g})$ , an element from one list can be compared with one or more elements from the other list. The mean directional difference between the two lists is the mean of the directional difference between the elements in the lists, that results in the lowest mean distance. We used the lowest possible distance, since we are interested in the optimal match. In handwriting recognition systems, recognition is often done at the end of gesture articulation [29, 13], and thus the best match approach chosen here is valid not only as an analysis measure, but could also be used in a working implementation of a recognition system.

To segment the articulated gesture into sub strokes, local minimums were found in the low-pass filtered speed signal of the gesture. This method is simple and is used in the handwriting literature [9].

Figure 7 shows the directional difference based on the 14 gestures in the experiment consisting of straight line segments only. There was no significant difference in this measure across the feedback conditions: $F_{3,24}$=2.28, $p$=0.105.

**Conclusions and Discussion of Experiment 1**

The main conclusions and the implications we drew from this systematic study of feedback on pen gesture reproduction are as follows.

First, the impact of auditory feedback on gesture reproduction was small. Even though we relaxed the requirement on the aesthetic auditory feedback and focused only on making the auditory feedback discriminable and responsive (instantaneous), the difference caused by auditory feedback in the shape aspects of the gestures produced was still statistically insignificant.

Second, although visual feedback had a significant impact on more aspects of the gestures reproduced, the magnitude of these impacts was also small except in terms of size and aperture of the closed gestures.

Third, the relatively small contribution of feedback, either visual or auditory, on gesture production supports open-loop theories of gesture control. Feedback appeared to be too slow for most aspects of gesture production. This is also consistent with the theory that handwriting relies on strong internal representations rather than feedback [28], notwithstanding the differences between natural handwriting and the digital gesture production in the experiment, including the amount of experience and the style of practice (distributed in natural handwriting vs. massed in our experiment).

Fourth, an important design implication from this experiment is that it is possible, in view of findings here, to minimize reliance on visual attention when designing the gesture set and recognition method*. As we can see, the overall shape, either in terms of length proportion (measure 3) or direction (measure 4) were not affected by the presence or absence of visual feedback. Only size and aperture were affected. The design of gesture interface should hence avoid using the latter two aspects as critical recognition information.

Fifth, there were a few unquantified benefits or impacts of auditory feedback in gesture production. During the experiment, some of the participants were observed to follow the rhythm of the sound with their head or body. In the conditions where no audio feedback was given, three of the nine participants indicated that they tried to recall the sound feedback in their memory while articulating the gestures. Three participants indicated that they felt more certain or more confident in producing the gestures with the auditory feedback than without. These comments were made in response to a general open-ended request of comments at the end experiment, or made unsolicited during the experiment.

## EXPERIMENT 2: THE JOY OF WRITING WITH MUSIC

In contrast to Experiment 1, which found some but relatively small impact of auditory feedback on gesture production, this experiment was focused on the aesthetic impact of auditory feedback. The goal was to find whether and how gesture input can be made more enjoyable, pleasant or stimulating through audio feedback.

### Experimental task and gesture recognizer development

To test the aesthetic of auditory feedback, we first needed to develop a pen gesture input task more practical than the task in Experiment 1. We chose to design and implement a revised version of EdgeWrite as our experimental system. EdgeWrite, developed by Wobbrock and Myers [27], is a text entry method designed to provide stability of motion for people with motor impairments. Its special alphabet leverages the physical edges for greater stability in text entry. A user enters letters by transversing the edges and diagonals of the square hole of a plastic fixture mounted on a PDA device. There were three reasons to choose EdgeWrite in this study. First, we needed to select an easy to learn but not necessarily high-performance gesture input method. One particularly inventive and impressive characteristic of the EdgeWrite alphabet is that although it differs in visual appearance from the Roman alphabet, the kinetic movement pattern is quite similar to the Roman alphabet, which makes EdgeWrite easy to learn [27]. Second, unlike Graffiti or Unistroke, the EdgeWrite alphabet consists of only straight lines, which eases the design of audio feedback in light of our ongoing research. Third, we wanted to test if reliable letter input could still be achieved without the physical edge fixture by designing and implementing a recognizer that takes advantage of the insight we have gained in the course of the current research project. For this reason we call our revised EdgeWrite system ELEW (Edgeless EdgeWrite) throughout the rest of the paper.

Unlike the original EdgeWrite [27], which recognizes each pattern according to the order in which the corners of the square hole are hit by the pen, the ELEW recognizer uses a proportional shape measure to determine which character (template) a gesture input was classified to. As shown in Experiment 1, proportional shape was invariant to visual feedback and hence may support heads-up use. Since the system does not require the pen gesture to reach exact locations (corners), we hypothesized that it could be reliable enough without the aid of the physical edges fixture in EdgeWrite. Furthermore, since ELEW does not rely on corners for recognition, we made some changes to the EdgeWrite alphabet to make the visual appearance of the ELEW alphabet more closely resemble the corresponding Roman character (See Figure 8).

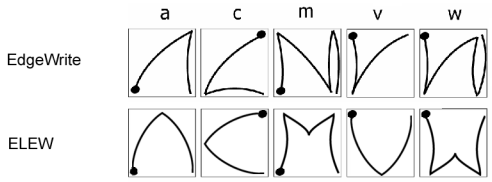


**Figure 8. Template pattern difference between EdgeWrite and ELEW.**

### Musical feedback

The experiment included four different auditory feedback conditions:

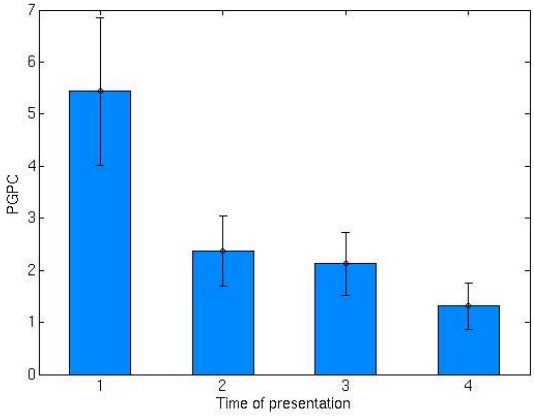


**Figure 9. Number of times help was used as a function of time.**

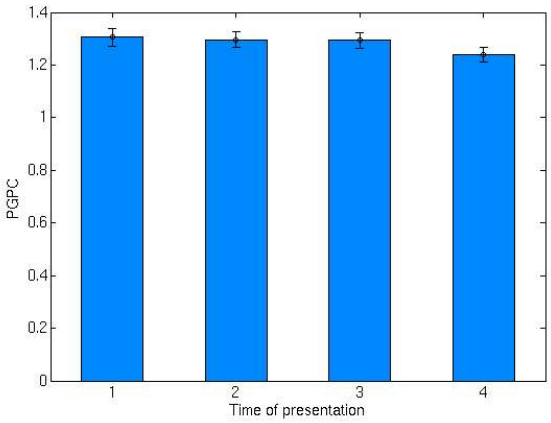


**Figure 10. PGPC as a function of time.**

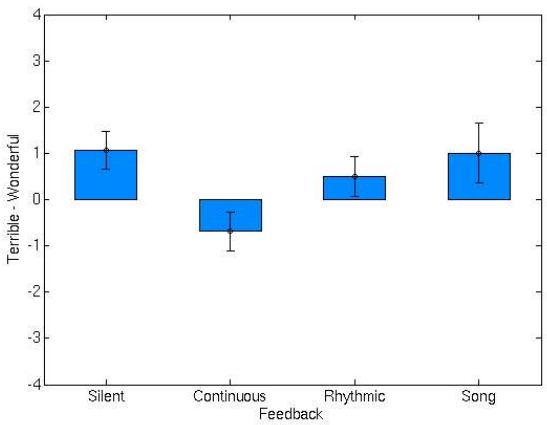


**Figure 11. Rating of each feedback condition on a Terrible-Wonderful scale. Mean and standard error mean values are shown for each feedback condition.**

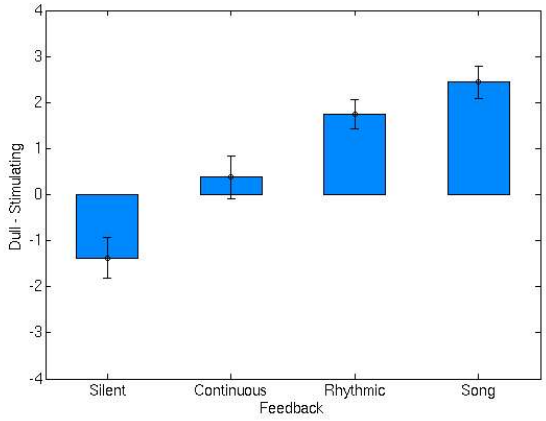


**Figure 12. Rating of each feedback condition on a Dull-Stimulating scale. Mean and standard error mean values are shown for each feedback condition.**

1. Silence
2. Continuous tone identical to the auditory feedback used in the previous experiment
3. Rhythmic feedback: Musical feedback based on guitar sounds and a drum loop
4. Song feedback: A song played at varying tempi determined by the users' average writing speed

The last two conditions can be called musical feedback. In general two directions can be taken in designing musical feedback. One is to compose the music algorithmically on-line based on component samples and synthesis models. The other is to use pre-recorded music in combination with meta-data describing the music, such as rhythm, key, and the structure of the musical piece. Using meta-data, it may be possible to transform the music during interaction while maintaining the original identity of the music. The advantage of this approach in comparison to pure algorithmic composition is that the user can choose the type of music as feedback. The user selects the songs that he/she wants, in the same way a musical piece is selected on the stereo or on an iPod. The music, however, will be transformed in real time based on the user's gesture input.

In this experiment, we designed Conditions 3 and 4 based on these two approaches, resulting in "rhythmic feedback" and "song playback" respectively. The rhythmic feedback was based on custom samples that were synchronized to a specific beat. The beat was defined by a drum loop and was played if and only if the user was writing. If no writing occurred for several seconds, the beat stopped and the system became silent. Whenever the user moved the pen in one of eight directions, a guitar sample was played. Each direction corresponded to a cord. The notes played in synchronization with the drum loop. This made the guitar sound blend nicely with the beat (Clip 3).

In the song playback condition, the same song, "Baya Baya" by the group Safri Duo, was played during writing, but the speed of playback was controlled by the average pen speed. If the user made a backspace stroke, the song played backwards until a new character was written. The scaling of the sound in time was done using linear interpolation of the waveform. One artifact of this method was that it also altered pitch with speed. On the other hand, this method was simple to use and the change of pitch made it easier to perceive difference when writing at different speeds (Clip 4).

### Experiment, evaluation and discussion

Sixteen participants, eight men and eight women, who had no training in music were selected, all between 20 and 50 years of age. Their prior experience with digital pen-based text input (mostly Graffiti) ranged from daily to none. The experiment task was to write the sentence "The quick brown fox jumps over the lazy dog" two times in each condition with ELEW. During writing the participant was asked to look at the screen where the only feedback was the recognized letters. A help screen showing the symbols for each letter in the alphabet could be displayed when the user pressed the space bar. The help screen was removed when the user started using the pen again. Before the experiment, a training session was given in which the participant wrote the alphabet from A to Z two times, and the testing sentence two times. There was no auditory feedback in the training session. Each participant participated in all feedback conditions. The order of conditions was balanced across participants.

Performance measures used included the number of times the help screen was used, task completion time, and pen gestures per character (PGPC). As one measure of efficiency, PGPC was defined as the ratio between the number of gestures made and the number of characters in the resulting string. It was calculated exactly the same way as KSPC (keystrokes per second) used

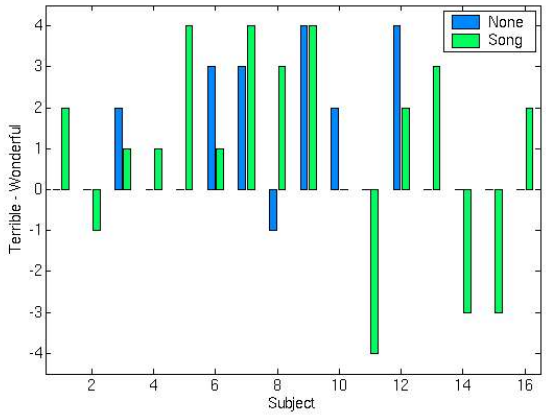


**Figure 13. Individual ratings of silent and song condition on the dimension "terrible (-4) – wonderful (+4)".**

by Wobbrock *et al.*, who adapted the term from stylus keyboarding studies [18] although in gesture interface there were no keystrokes per se. Note that to be comparable and consistent with the prior literature, back-space was not counted in the calculation [27].

The experimental results showed no significant difference in PGPC between any of the conditions: $F_{3,45} = 0.897$, $p$=0.45. The number of times the help screen was used was not significantly changed across conditions: $F_{3,45}$=0.938, p=0.43. Neither was the task completion time: $F_{3,45} = 0.283$, $p = 0.838$.

The number of times that the user had to call the alphabet symbols set over the course of the experiment decreased below 2 (out of a minimum 86 gestures required to write in each condition), showing that the ELEW alphabet was memorized easily (Figure 9). Completion time decreased to 140 seconds per "The quick brown ..." sentence, corresponding to 6.7 effective words per minute (WPM). PGPC decreased slightly in the course of the experiment to 1.28 in the last condition tested (Figure 10). Without the aid of the physical edges as in EdgeWrite and in less time taken than in Wobbrock *et al.*'s study [27], ELEW reached a very similar level of performance as EdgeWrite in terms of speed (WPM score) and accuracy (PGPC ratio), as reported in [27]. This level of performance achieved in an essentially heads-up condition in our experiment was also comparable to Sears and Arora's study of Graffiti and Jot [21] in which novice users wrote words while looking at the writing area (heads-down). The results showed that incorporating insights gained from Experiment 1 in the design of a gesture system (ELEW) could indeed be beneficial.

The joy and aesthetic aspects of the feedback conditions were evaluated using a modified version of the Questionnaire for User Interface Satisfaction [5]. The dimensions that were particularly relevant in the current context were: Terrible – Wonderful, Frustrating – Satisfying, Dull – Stimulating. Each scale was used by placing a mark on a scale with 9 levels. After the participants had completed the questionnaire, an open-ended interview was conducted with each participant.

On the dimension of "Terrible-Wonderful" (Figure 11), participants' ratings of the four sound conditions varied significantly ($F_{3, 45} = 2.853$, $p = 0.048$). Pairwise post hoc analysis showed that "silent" ($p = 0.009$) was significantly better rated than "continuous" condition. "song" was marginally better rated than "continuous" ($p = 0.078$), and "rhythmic" was not significantly different from "continuous" ($p = 0.133$).

On the dimension of "Frustrating-Satisfying", participants' ratings of the four sound conditions were not statistically significant ($F_{3, 45} = 1.285$, $p = 0.291$).

On the dimension of "Dull-Stimulating" (Figure 12), participants' ratings of the four sound conditions changed significantly ($F_{3, 45} = 22.571$, $p < 0.001$). Post hoc analysis showed that all pairwise comparisons were statistically significant ($p <= 0.016$), except between "rhythmic" and "song" (p=0.094). The "silent" condition was rated the most dull, "continuous" was slightly more stimulating, "rhythmic" was even more stimulating, and finally the "song" condition was rated the most stimulating.

Several participants commented on the "song" feedback. Some found it frustrating that the song speed increased as writing speed increased. Others liked it and thought it helped to stay in the "flow" of writing, and one participant wanted to keep writing to hear the whole song. Some commented on the song mapping, stating that they were not aware of what exactly was changed in the song during writing. Many participants further commented that they liked the continuous sound the least, but two participants thought that it would help them over longer use, even though they didn't like the sound. Some liked the rhythmic feedback better than the song feedback.

In summary, if and how well users like sound feedback for gesture production depends on the type of sound provided to them. In this experiment the participants did not like the continuous synthetic sound feedback, which was rated negatively on the dimension of "terrible – wonderful." The richer and more natural "rhythmic" and "song" conditions were much better liked in comparison.

Note that the best rated feedback, "song," was not judged better (or worse) than no sound at all ("silent") on the dimension of goodness (terrible – wonderful). This could be a result of the limited number of designs we have explored to date. There could potentially be still better designs to be developed. On the other hand, it is expected that there are situations and individuals that demand silence. Figure 13 shows the individual ratings of the silent vs. song condition, illustrating the personal preferences on the enjoyability of sound vs. no sound.

The experiment also unequivocally demonstrated that auditory feedback could make pen gesture production more exciting. All of the sound conditions, including the "continuous" condition, which was poorly rated on the dimension of "terrible – wonderful," were judged significantly more stimulating than the silent mode.

## CONCLUSIONS

Auditory information is a compelling resource to exploit in human computer interaction, particularly in view of the growing use of small mobile devices whose visual display space is limited. The fact that rhythm and patterns are easily perceived and memorized in sound makes auditory feedback an even more attractive modality for dynamic pen gestures which in essence are temporal-spatial patterns. Against this background, we conducted a series of studies on auditory feedback in pen-gesture input systems. Our initial iterative explorations showed that while it was easy to gain benefit from simple auditory feedback methods such as pronunciation of the gestured word, it was difficult to gain further advantage in gesture performance or learning time by employing more complex auditory feedback methods. To establish a baseline estimation, Experiment 1 formally evaluated gesture production accuracy as a function of auditory feedback as well as visual feedback. We found that while features such as the size of a gesture and the aperture of the closed gestures were influenced by visual or auditory feedback, other major features such as proportional shape distance were not affected by either. Although the experiment does not eliminate the possibility that more dramatic benefit could be found through broader explorations of the vast design space, the findings here support the theoretical hypothesis that any form of feedback is too slow for on-line gesture production which might be driven by a strong internal central representation [28]. There are important design implications of these findings. For example, instead of relying on visual or auditory feedback to close the aperture of closed gestures or more generally any features that are referential to previous segments of the gesture or other absolute positions, the design of gesture set and its recognizer should minimize this type of features to support heads-up use. Indeed, when we took this hint and designed our experimental gesture input system ELEW, whose recognizer used directional and shape information rather than positional information as employed in ELEW's predecessor EdgeWrite, our participants could quickly master ELEW without the aid of the physical bounding box used in EdgeWrite, and reached a performance level similar to that of EdgeWrite and Graffiti.

In addition to performance, emotional appeal and fun have recently being brought to the forefront of HCI research [19, 3]. It has been argued that a more aesthetic interface is also more usable [25]. Experiment 2 was focused on these aspects of auditory feedback in dynamic pen-gesture interfaces. It measured participants view of various subjective dimensions in four auditory feedback conditions when writing letters with ELEW: silent, continuous feedback, and two types of musical feedback. While these auditory conditions did not alter writing performance in terms of speed, learning or error rate, they were perceived very differently on the emotional dimensions. Participants rated the silent and song conditions more "wonderful," the silent condition more "dull," and the song condition most stimulating.

Overall, our exploration has laid a foundation for future design and research on auditory feedback in pen-gesture interfaces. It points out several general rules of auditory feedback: a few simple functions are obvious, gaining further performance and learning advantage is difficult, gesture set and its recognizer can be designed to minimize visual dependence, and emotional or aesthetic auditory effects are possible.

## ACKNOWLEDGMENTS

We thank members of the User group at IBM Almaden Research, and people from the HCI and music informatic groups at Department of Computer Science, University of Copenhagen. In particular Malcolm Slaney and Christopher Campbell for insightful discussions and comments.

## REFERENCES

1. Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. The CIPIC HRTF Database, *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics* (2001), 99-102.

2. Blattner, M., Sumikawa, D., and Greenberg, R. Earcons and icons: Their structure and common design principles. *Human Computer Interaction*, 16 (1990), 523-531.

3. Blythe, M., Monk, A., Overbeeke, C., and Wright. P.C. (Eds.) Funology: From Usability to User Enjoyment. Dordrecht: Kluwer, 2003.

4. Buxton, W. Speech, Language & Audition. Chapter 8 in Baecker, R.M., Grudin, J., Buxton, W., and Greenberg, S. (Eds.). *Readings in Human Computer Interaction.* San Francisco: Morgan Kaufmann Publishers, 1995.

5. Chin, J.P., Diehl, V.A., and Norman, K.L., Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. *Proceedings of CHI* (1988), 213-218.

6. Clarke, E.F., Rhythm and Timing in Music. Chapter 13 in Deutch, D. (Ed.). *The Psychology of Music* (Sec.ed.). Academic Press, 1999.

7. Deatherage, B. H. Auditory and Other Sensory Forms of Information Presentation. In H. P. Van Cott & R. G. Kinkade (Eds), *Human Engineering Guide to Equipment Design* (Revised Edition). Washington: U.S. Government Printing Office, 1972.

8. van Doorn, R., and Keuss P. The role of vision in the temporal and spatial control of handwriting. *Acta Psychologica*, 81 (1992), 269-286.

9. van Doorn, R., and Keuss P. Does production of letter strokes in handwriting benefit from vision? *Acta Psychologica*, 82 (1993), 275-290.

10. Fletcher, H., Blackham, E.D., and Stratton, R. Quality of piano tones, *Journal of the Acoustical Society of America,* 34 (6) (1962), 749-761.

11. Gaver, W. W. The SonicFinder: An Interface that Uses Auditory Icons. *Human Computer Interaction*, 4(1) (1989), 67-94.

12. Ghez, C. Rikakis, T. DuBois R.L. and Cook P.R. An Auditory Display System for Aiding Interjoint Coordination. *Proceedings of ICAD* (2000).

13. Goldberg D., and Richardson C. Touch-Typing with a Stylus. *Proceedings of InterCHI* (1993), 80-87.

14. Kristensson, P-O., Zhai, S., SHARK$^2$: A Large Vocabulary Shorthand Writing System for Pen-based Computers, *Proceedings of UIST* (2004).

15. Kurtenbach, G., and Buxton, W. User Learning and Performance with Marking Menus. *Proceedings of CHI* (1994).

16. Legge, D., Steinberg, H., and Summerfield A. Simple measures of handwriting as indices of drug effects. *Perceptual and Motor Skills*, 18 (1964), 549-558.

17. Loeb, R.G. and Fitch, W. T. A Laboratory Evaluation of an Auditory Display Designed to Enhance Intraoperative Monitoring, *Anesth Analg,* 94 (2002), 362-368.

18. MacKenzie, I.S. KSPC (Keystrokes per Character) as a Characteristic of Text Entry Techniques. *Proceedings of Mobile HCI*, Springer-Verlag (2002), 195-210.

19. Norman, D.A. Emotional Design: why we love (or hate) everyday things. New York: Basic Books, 2004.

20. Plomp, R. The ear as a frequency analyzer. *Journal of the Acoustical Society of America*, 36 (1964), 1628-1636.

21. Sears, A., Arora, R. Data entry for mobile devices: An empirical comparison of novice performance with Jot and Graffiti. *Interacting with Computers,* 14 (2002), 413-433.

22. Singh, P.G. Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre? *Journal of the Acoustical Society of America,* 82 (1987), 886-899.

23. Smyth, M. and Silvers, G. Functions of vision in the control of handwriting. *Acta Psychologica*, 66 (1987), 47-64.

24. Teulings, H.L., and Schomaker, L., Invariant properties between stroke features in handwriting. *Acta Psychologica*, 82 (1993), 69-88.

25. Tractinsky, N., Katz, A. S., Ikar, D. What is beautiful is usable. *Interacting with Computers*, 13 (2000), 127-145.

26. Welch, R.B. Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. In: *Cognitive Contributions to the Perception of Spatial and Temporal Events*, Aschersleben, G., Bachmann, T., and Musseler, J. Amsterdam (eds): Elsevier, 1999.

27. Wobbrock, J.O., Myers, B.A., and Kembel, J.A. EdgeWrite: A Stylus-Based Text Entry Method Designed for High Accuracy and Stability of Motion. *Proceedings of UIST* (2003), 61-70.

28. Wright, C.E. Generalized motor programs: Reexamining claims of effector independence in writing. In M. Jeannerod (Ed.), *Attention and performance XIII* (pp. 294-320), 1990.

29. Zhai, S., and Kristensson, P.-O. Shorthand Writing on Stylus Keyboard. *Proceedings of CHI* (2003), 97-104.