# Applying Sonification to Sketching in the Air With Mobile AR Devices

Haonan Xu, Fei Lyu, Jin Huang, and Huawei Tu

***Abstract*—With more and more mobile devices (such as smart phones and tablets) supporting augmented reality (AR), using these devices to sketch in mid-air has become a popular application direction. However, due to the small display size of these devices, the user's field of view is limited and cannot see the context of graphics, which leads to the drawn graphics deviating from their expectations. In this article, we applied sonification technology to mid-air sketching with mobile AR device, and proposed a new method to address this problem. In our first experiment, we verified the feasibility of our method. Our experimental results showed that sonification can effectively reduce the deviation caused by a narrow field of view. In our second experiment, we further explored the application ability of this method in a wider range of sketching. Our experimental results showed that sonification can improve the aspect ratio of the drawn graphics. In addition, the results of the NASA-TLX questionnaire showed that the participants' mental demand and effort decreased significantly, and their subjective performance increased significantly. We proposed a new method, which can effectively improve the accuracy of mid-air sketching with a mobile AR device.**

***Index Terms*—Augmented reality (AR), mid-air sketching, mobile applications, narrow field of view (FOV), sonification.**

## I. Introduction

SKETCHING is an important and commonly used creative expression method. Compared with the traditional drawing method, sketching with augmented reality (AR) devices is a novel way. It allows designers to create the conceptual design in real-world proportion [20], combine the design with the real world [31], and complete rapid evaluation [32]. With more and more mobile phones supporting AR in recent years, we can draw graphics in mid-air directly using mobile phones. Before that, mid-air sketching can only be applied by complex and expensive virtual reality (VR) equipment and head-mounted AR device, such as Microsoft HoloLens. However, mid-air sketching with a mobile AR device has a problem, that is, the user's field of view (FOV) is narrow due to the small size of the display device. When the graphic size exceeds the size of the display device, it is difficult for users to see the context, resulting in a large drawing deviation [1]. This phenomenon is particularly evident in mobile AR devices (see Fig. 1).

Manuscript received 11 October 2021; revised 10 February 2022 and 24 May 2022; accepted 8 June 2022. Date of publication 7 July 2022; date of current version 15 November 2022. This work was supported by the National Natural Science Foundation of China under Grant 61972045. This article was recommended by Associate Editor V. Fuccella. *(Corresponding author: Fei Lyu.)*

Haonan Xu and Fei Lyu are with the School of Digital Media and Design Arts, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: xuhaonan42@bupt.edu.cn; feilyu@hotmail.com).

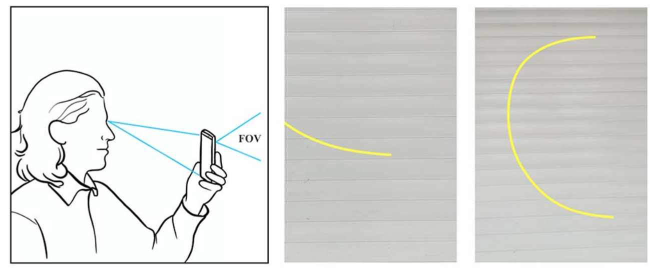
Jin Huang is with the Beijing Key Laboratory of Human-Computer Interaction State Key Laboratory of Computer Science, Institute of Software Chinese Academy of Sciences, Beijing 100190, China (e-mail: hj34130226@126.com).

Huawei Tu is with the Department of Computer Science and Information Technology, La Trobe University, Bundoora, VIC 3086, Australia (e-mail: h.tu@latrobe.edu.au).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by School of Digital Media and Design Arts.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/THMS.2022.3186592.

Digital Object Identifier 10.1109/THMS.2022.3186592




(a) (b) (c)

Fig. 1. (a) Narrow FOV of mobile AR device. (b) User cannot see the context when drawing large-scale graphics. (c) User can only see the whole graphic by stepping back.

At present, only a few studies have focused on the problem of narrow FOV of mobile AR devices. Mobi3dsketch [1] is a three-dimensional (3-D) AR drawing system, which allows users to observe the context by moving backward away from the graphics, and realizes long-distance drawing through snapping points or plane proxy. New strokes will be mapped to the snapping points or the selected proxy. However, this article points out that users think that the long-distance projection makes the strokes deviate from their expectations. SymbiosisSketch [20] is a hybrid AR drawing system, which allows users to reduce large-size graphics or 3-D models *in situ* before drawing. But due to the inevitable hand jitter, the error will be synchronously amplified when the lines drawn on a small scale are enlarged. If users want to use mobile AR devices for more accurate and free drawing, they need to ensure the accuracy of strokes drawn in the original size.

One solution to the problem is to supplement additional information through auditory, tactile or multimodal feedback. Sound is an important way of information transmission, so using auditory feedback to compensate for visual information is one direction of the HCI field. The most typical applications are to assist the blind, such as indoor navigation [35], or to identify surrounding object information [34]. Sonification is one of the auditory feedbacks. It is a way of transmitting information using nonspeech sound. It is the corresponding concept of visualization. At present, some studies have applied sonification to navigation without visual information [9]–[11]. Therefore, we believe that it is effective to use sonification to alleviate the problem of narrow FOV of mobile AR device.

In this article, we apply sonification to mid-air AR sketching to provide a new solution to the problem of narrow FOV. We map the pitch of the French horn and Cello to the *X*-axis and *Y*-axis of a 2-D plane. During the drawing process, as the coordinates of the *X*-axis and *Y*-axis increase or decrease, the pitch of the two sounds will rise or fall. By repeating practice, participants can establish a corresponding graphic-sound model, so they can perceive the direction and length of strokes according to the pitch change. We verified this hypothesis and conducted a practical exploration. Our first experimental results showed that sonification can effectively increase the drawing accuracy. Compared with the no sound condition, the deviation of the graphics decreased by 28.5% in the sonification condition. The second experiment found that the aspect ratio of the graphics drawn by the participants was more accurate under the sonification condition. Also, NASA-LTX questionnaire showed that in sonification conditions, mental demand and effort decreased significantly, and subjective performance increased significantly.

We propose the use of sonification to alleviate the problem of narrow FOV of mid-air AR sketching. This is the first study to use sonification to address this kind of problem. Our results showed that sonification was an effective way of information transmission and could improve the quality of graphics in mid-air AR sketching.

## II. Related Work

In the first part, we point out the definition of sonification, several forms of sonification, and its application in our life. The second part summarizes the related work of graphic sonification. The third part discusses the recent research on mid-air AR sketching and the existing problems.

### A. Sonification

*Sonification* refers to the way of using nonspeech sounds to convey information. In particular, sonification conversions logical relationships in data into perceptual relationships in sound [2]. Sonification is the auditory counterpart of visualization. The main purpose of sonification is to optimize communication efficiency [3], or to supplement and even replace visual feedback in some scenarios [17].

The Sonification Handbook [4] introduces several methods of signification. At present, it mainly includes the following methods: audification, auditory icon, earcon, parameter mapping sonification, and model-based sonification. *Parameter mapping sonification* means to create a set of mapping relationships between data and sound. The properties of the sound (such as volume, pitch, spatial position, etc.) change synchronously with changes in the data. This is a simple and effective method of sonification and becomes the most widely used sonification method.

The most familiar sonification application in our life should be the Reverse Parking Sensors and Geiger Counter. In academic research, sonification has been applied to sports [5], [6], health [7], [8], navigation [9]–[11], handicraft [12], art [13], [14], and many other fields.

### B. Graphic Sonification

Sound Graphs system [15] was used to sonify simple graphics. In the data-sound mapping relationship, the *Y*-axis of the graphics was mapped to the frequency of a sound, and the sound was played sequentially as the *X*-axis increased. Users could perceive the distribution of the *Y*-axis on the *X*-axis. For simple graphics, this is an intuitive sonification method. Kim *et al.* [16] used a similar design to help blind people perceive mathematical graphics. Furthermore, they applied different timbres to represent the four quadrants, as well as frequency and speech to express differentiability. A system designed by Harada *et al.* [17] was used to sonify spatial information that composed of eight radial directions. A cursor traces the path at a constant speed, and the tangent direction of the cursor position being continuously sonified. The results found that users can understand the directional sound mapping method in a short time and distinguish different path shapes correctly. One image-to-sound conversion system designed by Meijer [18] could sonify an image of 64∗64 pixels with 16 gray levels. The system scans images from left to right, one column of pixels at a time. The frequency, duration, and brightness of the sound were mapped to the *Y*-axis, *X*-axis, and the gray level of each pixel. Rigas and Alty [19] divided the graphic display area into 40∗40 grids. Two pure notes, organ and piano, were used for navigation. The piano notes always appear before the organ. After a short learning period, users could recognize different graphics of different sizes and positions through the sound sequence.

### C. Mid-Air Sketching in AR

The mobile AR sketching system Mobi3Dsketch developed by Kwan and Fu [1], has been used to create 3-D conceptual designs. Mobi3Dsketch allowed users to stretch the curve into a curved surface, create a plane quickly according to the posture of the mobile device, and copy and rotate virtual contents. In addition, the user could observe the whole graphics by stepping back, and use the snapping point or plane proxy for long-distance drawing. The new stroke will be mapped to the selected point or proxy through projection. SymbiosisSketch [20] was a hybrid sketching system consisting of Microsoft HoloLens, a 3-D motion-tracked stylus and a tablet. Users could use the stylus to draw 3-D strokes in mid-air, and generate a surface from the stroke. The stroke drawn on the digital tablet could be mapped to the 3-D surface in mid-air. For large-scale graphics, users could use the zoom function to reduce the graphics before drawing. SymbiosisSketch combined the advantages of 2-D sketching and 3-D sketching to help users quickly create designs of any scale. ARPen [33] was a lightweight AR sketching system, which is composed of an iPhone and a 3-D printing pen. There was a wireless button at the tip of the pen and a visual marker at the end of the pen. The system used ARKit to track the position of the mobile phone in space, and tracked the visual marker for the position of the pen. The button at the tip of the pen was used to start or end drawing. In this way, users can draw 3-D graphics in the AR program and preview them on mobile phone. In sketching, alignment is a very important operation, which helps us create content more in line with aesthetic and design principles. SnapToReality [31] is an alignment technology. By identifying and extracting the edges of physical objects, it can align the virtual content with the physical objects in the real world. SnapToReality also supports stretching, rotating and other operations on virtual content. Kratz and Rohs [37] presented a $3 Gesture Recognizer, which can recognize user's gestures using data from the input device's 3-D accelerator. The $3 Gesture Recognizer has the advantages of low cost, fast deployment, and good scalability. In particular, it can recognize real 3-D gestures, not just 2-D gestures in 3-D space. Therefore, it has a high potential for drawing prototypes in AR. A course organized by LaViola and Keefe [36] systematically introduced the development of the 3-D spatial interface. In addition to introducing the common tasks of 3-D user interface and corresponding solutions, the course also elaborates the research history and latest progress of 3-D user interface for art and design, as well as some application cases.

Our review indicates that only a few studies have focused on the problem of narrow FOV of mid-air AR sketching. Mobi3dsketch and SymbiosisSketch provide long-distance drawing and scaling functions, but still cannot improve the accuracy of users' sketching. We believe that if we want to use mobile AR devices for more accurate and free mid-air sketching, we need to ensure the accuracy of strokes drawn by users in the original size. Sonification is worth trying because it transmits information through auditory channels. In addition, due to the strong correlation between the auditory system and the motor system, it has high application potential in the scene of perceiving own actions [24]. Therefore, this study will apply sonification to mid-air AR sketching to explore whether sonification can improve the quality of AR drawing with mobile AR devices.

## III. Research Questions and Objectives

In order to verify whether sonification can improve the accuracy of mid-air AR sketching and explore the effects and problems in the application of sonification, two user experiments were carried out in this study. We aimed to address the following three research questions:

1) Whether sonification can improve the accuracy and speed of mid-air sketching with a mobile AR device?
2) Would different sound parameters and graphic sizes affect the accuracy and speed?
3) Whether signification has positive effects on copying tasks in sketching?

Experiment one discussed the first and second questions, and experiment two explored the third question.

While drawing 3-D graphics in AR is an exciting way to create, 2-D graphics are equally widely used in 3-D space, and that 2-D graphics are easier to draw. Therefore, in our work, we target 2-D graphics in AR space.

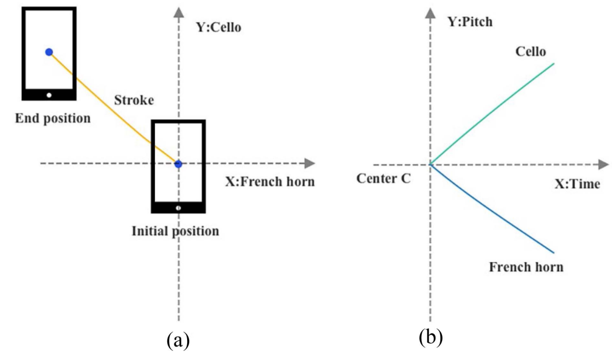



Fig. 2. (a) Mobile phone moves from the initial position to the end position. The orange line is the moving path, that is, the drawn stroke. (b) Change of pitch during stroke drawing in (a).

## IV. Experiment One

In the first experiment, we wanted to verify whether sonification can improve the accuracy and speed of mid-air AR sketching. In addition, we were also interested in the effects of different sound parameters and different graphic sizes on drawing accuracy. In this experiment, participants regarded the mobile phone as a brush and drew graphics in the air. In the process of drawing, the system detected the motion direction and distance of the mobile phone through the motion tracking algorithm, mapped them to the change of pitch, and transmitted it to the participants through headphones. During the drawing process, the movement of the mobile phone was continuous, so the change of tone was continuous too (as shown in Fig. 2).

### A. Experimental Design

*1) System Description:* We chose the parameter mapping sonification to map the pitch of the French horn to the *X*-axis and the pitch of the Cello to the *Y*-axis in a 2-D plane in the air. The center of the plane was the coordinate origin (see Fig. 2). We took central C as the pitch of the coordinate origin. When the stroke moves to the right or up, the pitch rises, and moves to the left or down, the pitch falls. Due to the lack of guidance on timbre selection, relevant studies only require sufficient timbre discrimination [4]. We chose the French horn and the Cello because their timbre is obviously different. The two sounds were sampled from GarageBand [40], an audio software developed by Apple. Pitch was selected as the target attribute because it is the most widely used attribute in sonification studies [21]. In addition, our recognition of sound is disturbed by other sounds [11]. The more sounds we hear at the same time, the more difficult it is to recognize. In order to reduce the impact of sound on participants, we have limited the volume of these two sounds. According to the speed component $V_x$ in the horizontal and $V_y$ in the vertical of the equipment at the current time, the volume of the French horn $Vol_x$ and the volume of the Cello $Vol_y$ meet

TABLE I
FOUR SOUND LEVELS AND THREE GRAPHIC SIZES IN EXPERIMENT ONE

| Variable | | Level |
|---|---|---|
| Sound | No Sound | No Sound |
| | Sonification | 5Cent/cm |
| | | 10Cent/cm |
| | | 20Cent/cm |
| Graphic Size | | 30cm |
| | | 60cm |
| | | 90cm |

the following requirements:

$$\mathrm{Vol_x} = \sin\left(\frac{\pi}{2} \times \frac{\mathrm{V_x}}{\mathrm{V_x} + \mathrm{V_y}}\right) \times 100\% \tag{1}$$

$$\mathrm{Vol_y} = \sin\left(\frac{\pi}{2} \times \frac{\mathrm{V_y}}{\mathrm{V_x} + \mathrm{V_y}}\right) \times 100\%. \tag{2}$$

We projected the graphics drawn by participants onto the 2-D plane to avoid the influence of depth on the experiment. Although depth is an important dimension in AR drawing, we also need to draw a large number of 2-D graphics even in 3-D space. The plane was located about 40 cm in front of the participant, and the center of the plane was at the same height as the center of the screen when the program started. Therefore, users can restart the program to reset the height of the plane and complete the experiment with a comfortable posture.

We used the AR software development kit ARCore [41] developed by Google to realize the AR function of the program. ARCore can accurately detect the posture and position of the mobile phone. When the program starts, the position of the mobile phone was initialized as the origin of the world coordinate system ($W_o$), the virtual pen tip ($P_o$) was located 40 cm in front of $W_o$, and the mapping relationship between the world coordinate system ($W$) and the drawing plane coordinate system ($P$) was established. During the drawing process, we used ARCore to update the phone's spatial position ($W_i$) and the virtual pen tip's spatial position ($P_i$) in real time. Note that the Z-axis of $P_i$ was fixed, that is, $P_{i \cdot z} = P_{o \cdot z}$ .

*2) Experimental Variables:* The experiment was a within-subject design with two independent variables (see Table I). The first variable was sound, which contains two conditions: no sound and sonification. In order to further study the effects of different sound parameters on the experiment, we further divided sonification into three-pitch change rate levels: 5, 10, and 20 cent/cm. The second variable was the graphic size, which contains three levels: 30, 60, and 90 cm. Cent is a unit used to express pitch difference in music theory. The pitch rises with the frequency increase, but the pitch and frequency are not linearly related. For every 1200 cent increase, the pitch rises by an octave. We use cent to ensure that the degrees of pitch change is the same when moving the same distance. Cent/cm refers to the rate at which the pitch changes with the moving distance. When moving the same distance, the bigger the pitch change rate, the more pitch rise (or fall), and the easier it is for users to recognize.

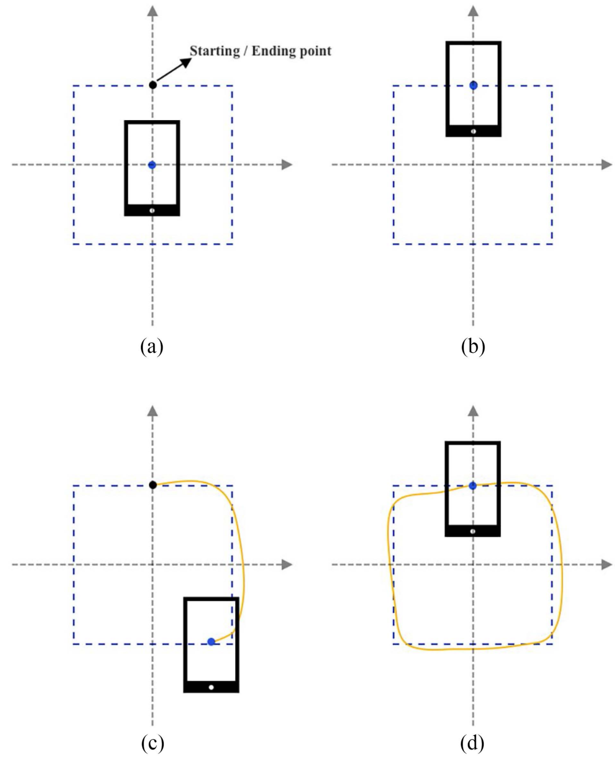



Fig. 3. Diagrammatic sketch of the stroke task. The axes were not displayed in the dictation task. (a) System was initialized. (b) Participants controlled the blue dot to touch the grey dot and the task began. (c) Participants controlled the phone along the edge of the reference graphic. The yellow line was the path. (d) Back to the start point and touched to the grey dot again, this task was over.

This concept refers to the pixel density unit pixels per inch (PPI) of visual display. The higher the PPI, the higher the definition of the display.

### *B. Task and Procedure*

Similar to previous work [38], when evaluating a new drawing system, we asked participants to draw some simple graphics, and took the indicators of these graphics as the dimensions of evaluating the system. We chose a square as the reference graphic in this experiment.

In the experimental program, a blue dot fixed in the center of the screen was used as the cursor to indicate the position where the stroke appeared. The cursor moved with the mobile device, and the moving track was the graphic drawn by the participants. A gray dot, fixed to the top of the graphic, was used as the controller to control the start and end of the task. When the blue dot first touched the gray dot, the task started. When the blue dot touched the gray dot again, the current task ended and the system waited for the participant to start the next task (as shown in Fig. 3). Therefore, throughout the experiment, participants did not need to touch the screen with their fingers. In the interval between the end of the current task and the start of the next task, participants could take a rest, and the rest time depended on themselves.

In this work, we used sound feedback to help participants draw graphics. However, it was not easy to understand and distinguish


(a) (b) (c) (d)

Fig. 4. Screenshots of experiment one. The text in the lower left corner of the screen was the current progress. The light-blue line was the reference graphic, the light-yellow line was the graphic drawn by the participant, and the gray line in the middle was the coordinate axis. (a) Experimental program started. (b) Drawn the reference graphic in stroke task. (c) Completed a stroke task, but did not start the next task. (d) In the dictation task, the axes and reference graphic were hidden.

these sound feedbacks, and the short-term memory formed by these sounds was easy to be forgotten. Therefore, we designed the practice task before the participants drew those graphics. The practice task did not take place before the whole experiment, but before the drawing task, so as to ensure that the participants had a clear memory of the sound feedback in the current condition.

The practice task was the stroke task, and the drawing task was the dictation task. In the stroke task, the reference graphic that participants need to draw was displayed in AR. Participants drew along the outline of the reference graphic, understood and remembered the sound feedback in the current condition. In the dictation task, the reference graphic was hidden, and participants needed to draw the reference graphic as accurately as possible in the same condition. In each experimental condition, participants needed to complete 3 stroke tasks and 1 dictation task in sequence. This cycle was repeated in the next experimental condition until the end of the experiment. Fig. 4 shows screenshots of the experimental procedure at different stages.

In the pilot study, we asked participants to listen as carefully as possible to the sound feedback during the drawing process and to draw more accurate graphics based on the sound feedback. We found that the completion time of all successfully completed tasks was greater than 3 s, while the completion time of tasks that ended early due to operational errors was generally less than 3 s. Therefore, in the formal experiment, if the completion time of any task was less than 3 s, the task would be regarded as invalid, and the system would automatically return to the place where the last task ended. Under each experimental condition, 3 stroke tasks and 1 dictation task were regarded as 1 trail. There were 12 conditions in this experiment, that is, 12 trials to be completed, which was 1 block. Each participant needed to repeat 4 blocks to generate enough data and balance the learning effect. The experimental conditions in each block appeared randomly. Each participant needed to complete 4 blocks $\times$ 12 trials $\times$ (3 stroke task + 1 dictation task) = 192 drawing tasks, which took about 40 min.

The researchers first stated the form and purpose of the experiment to the participants and demonstrated the operation and process of the experiment to the participants using the experimental equipment. Before the formal experiment began, participants spent about 15 min to understood the system and related operations.

### *C. Participants and Apparatus*

Thirteen students were recruited for this experiment. Five males and eight females, aged 21–31. Eleven participants were right-handed and the remaining two participants were left-handed. Eight participants had more than one year of hand-painting experience, four participants had less than half a year of hand-painting experience, and one participant had no hand-painting experience at all. Ten participants had prior experience in using AR applications (not drawing applications), while the other three participants did not.

The experimental program was based on ARCore and developed by Unity. A mobile phone (Realme X2) equipped with Android 10 version system was used to run the program. A pair of headphones with passive noise-canceling (Audio-technica CKB50) was chosen to transmit sound and isolate ambient noise. Each participant signed informed consent. Before the experiment, the researchers gave a brief introduction to the experiment.

### *D. Measurements*

Completion Time: The time for participants to complete one dictation task.

Deviation: The average deviation between the sample point of the drawn square and the corresponding sample point of the reference square [22]. The deviation can effectively calculate the distance between the two graphics and measure the accuracy of participants' reproduction. Each square had 96 sample points (multiple of 4 sides), evenly distributed on four sides.

The refresh rate of the system was 30 Hz. After participants finished the dictation task, the perimeter ($C$) of the drawn graphic was obtained, and the average distance ($L$) between 96 sampling points was calculated. Define the starting point as $P_0$ , increase L, and take the point closest to the landing point as $P_1$, and so on until the graphic was sampled (as shown in Fig. 5)

$$\text{Deviation} = \frac{1}{\text{k}} \sum_{\text{n,m} = 1}^{\text{k}} \sqrt{\left(\text{x}_{\text{P}_\text{n}} - \text{x}_{\text{P}_\text{m}}\right)^2 + \left(\text{y}_{\text{P}_\text{n}} - \text{y}_{\text{P}_\text{m}}\right)^2} \quad (3)$$

### *E. Results*

The data did not pass the Shapiro Wilk normality test ($p < 0.05$), so Friedman test was used for data analysis.

*1) Completion Time:* Four sound conditions had no significant effects on the completion time. The graphic size had a significant effect on the completion time, $\chi^2$ (2) = 390.58, $p < 0.01$ (see Fig. 6). There were significant differences between 30 cm (*Median* = 5.52 s), 60 cm (*Median* = 7.77 s) and 90 cm (*Median* = 9.85 s). The larger the graphic size, the longer

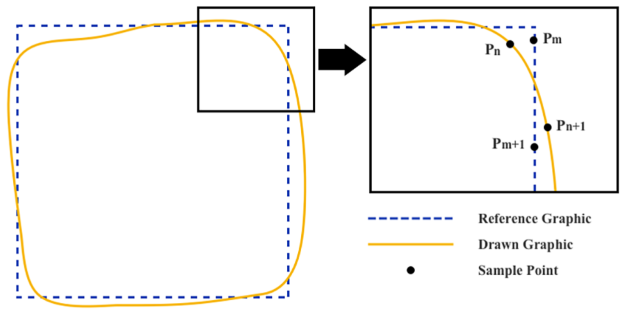



Fig. 5. $P_n$ and $P_m$ represented the sample point from the reference graphic and drawn graphic. $x_{P_n}$ and $y_{P_n}$ were the *X*-axis and *Y*-axis coordinate values of point $P_n$. The distance between each corresponding point represents the deviation.

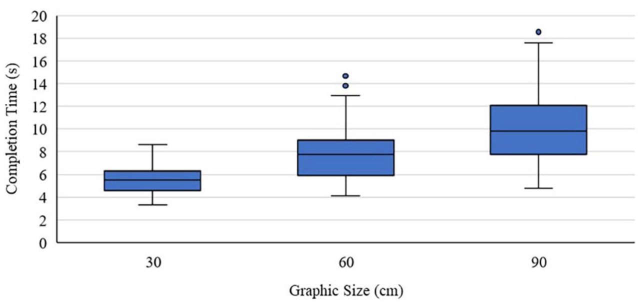



Fig. 6. Completion time of different sizes of graphics. Outliers based on the interquartile range (IQR) rule.

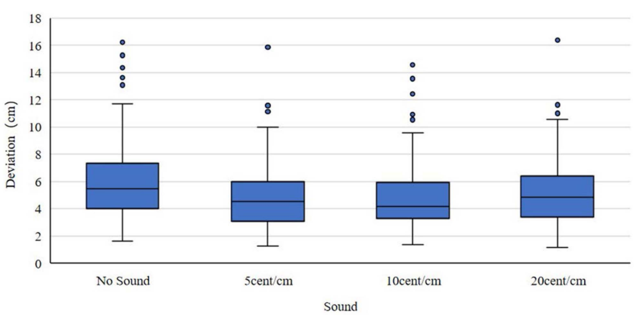



Fig. 7. Effect of different sound condition on deviation. Outliers based on the IQR rule.

the completion time. In the same graphics size, four sound conditions had no significant difference in the completion time.

*2) Deviation:* Sound conditions had a significant main effect on the deviation, $\chi^2$ (3) = 20.81, $p < 0.01$ (see Fig. 7). No sound (*Median* = 5.45 cm) was significantly higher than 5 cent/cm (*Median* = 4.56 cm), 10 cent/cm (*Median* = 4.19 cm) and 20 cent/cm (*Median* = 4.83 cm). But the deviation had no significant difference in 3 pitch change rate levels.

Significant effects were observed for all three graphic sizes, $\chi^2$ (2) = 101.38, $p < 0.01$ (see Fig. 8). The 90 cm (*Median* = 5.83 cm) resulted the highest deviation followed by 60 cm (*Median* = 4.87 cm) and 30 cm (*Median* = 3.49 cm).

The effect of the combination of graphic size and sound variable on the deviation was checked. When the graphic size was 30 cm, the sound variable had a significant effect on the

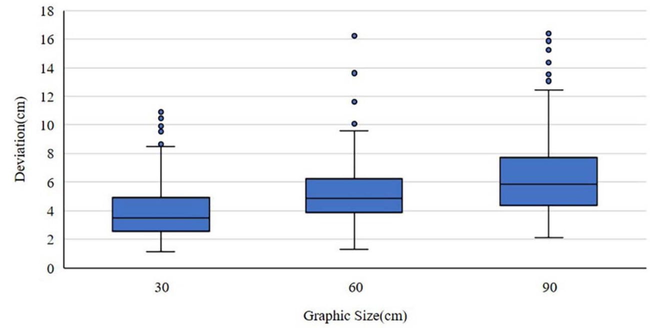



Fig. 8. Effect of graphic sizes on deviation. Outliers based on the IQR rule.

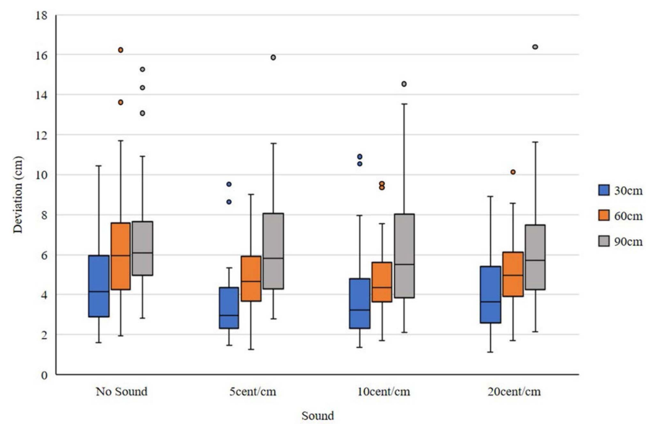



Fig. 9. Deviation in different of sound levels and graphic sizes. Outliers based on the IQR rule.

deviation, $\chi^2$ (3) = 13.98, $p < 0.05$. No sound (*Median* = 4.17 cm) was significantly higher than 5 cent/cm (*Median* = 2.98 cm) and 10cent/cm (*Median* = 3.22 cm). When the graphic size was 60 cm, no sound (*Median* = 5.95 cm) was significantly higher than the 5 cent/cm (*Median* = 4.66 cm) and the 10 cent/cm (*Median* = 4.36 cm), $p < 0.05$. However, when the graphic size was 90 cm, the sound variable had no effect on the deviation.

Fig. 9 shows the trend of deviation in different sound conditions. When the graphic size was 30 cm, the deviation was reduced by 28.5% at 5 cent/cm. When the graphic size was 60 cm, the deviation was the lowest at 10 cent/cm, which was reduced by 26.8%.

The effect of the experimental stage on the deviation was analyzed to check whether the participants' performance declined over the process of the experiment. In all conditions, the deviation had no significant differences in the four experimental stages (see Table II).

## V. Experiment Two

The results of experiment one showed that with sonification, the accuracy of simple graphics was indeed improved, but the speed did not change. However, in practice, the graphics drawn by users are more complicated, and there is no repeated practice before drawing. Therefore, users can only improve the drawing accuracy by modifying the path according to the real-time auditory feedback. Therefore, we conducted experiment two to

TABLE II
EFFECT OF EXPERIMENTAL STAGE ON THE DEVIATION

| Size | Sound | Median of Deviation | | | | p-value |
|---|---|---|---|---|---|---|
| | | Block1 | Block2 | Block3 | Block4 | |
| 30cm | No Sound | 3.72 | 4.59 | 3.99 | 3.93 | 0.45 |
| | 5cent/cm | 3.96 | 3.45 | 2.97 | 2.64 | 0.12 |
| | 10cent/cm | 3.18 | 3.23 | 3.51 | 3.20 | 0.97 |
| | 20cent/cm | 3.87 | 3.66 | 3.20 | 3.31 | 0.66 |
| 60cm | No Sound | 6.09 | 5.49 | 5.99 | 6.31 | 0.94 |
| | 5cent/cm | 4.57 | 4.55 | 5.05 | 4.81 | 0.44 |
| | 10cent/cm | 4.15 | 4.35 | 4.68 | 3.87 | 0.72 |
| | 20cent/cm | 4.83 | 4.87 | 4.82 | 5.25 | 0.97 |
| 90cm | No Sound | 5.80 | 6.52 | 6.37 | 5.66 | 0.44 |
| | 5cent/cm | 6.02 | 5.85 | 5.71 | 5.75 | 0.81 |
| | 10cent/cm | 5.17 | 4.29 | 5.73 | 6.22 | 0.15 |
| | 20cent/cm | 5.41 | 6.84 | 5.62 | 4.74 | 0.41 |

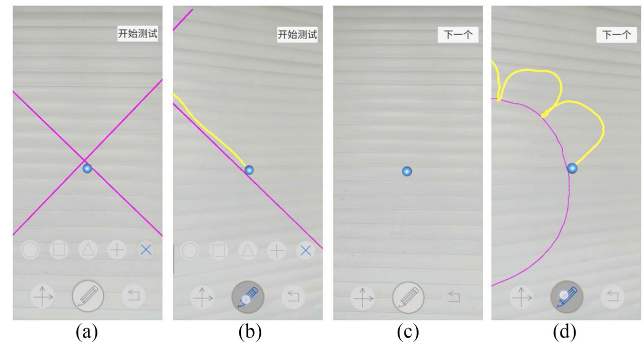



Fig. 11. Screenshots of experiment two. The button at the top right of the screen was used to control the start and next step of the experiment. The three buttons from left to right at the bottom of the screen were: reset coordinate center, draw, and undo. (a) In the practice experiment, participants could switch the practice graphics by clicking the button on the penultimate row of the screen. (b) In the practice experiment, participants could long press the draw button to draw practice graphics. (c) Entered the formal experiment without drawing any graphics. (d) Drew the “sunflower” in the formal experiment.

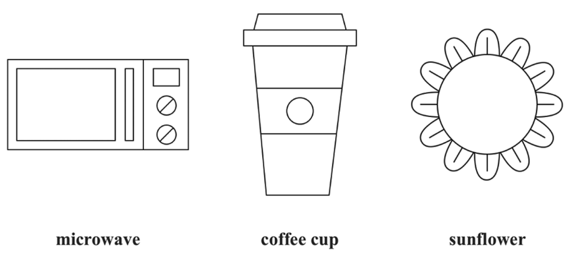



Fig. 10. Reference graphics to copy.

investigate the effects of sonification on performing copy tasks. Besides, we wanted to know the effect of sonification on the task load.

## A. Experimental Design

Experiment two was also designed as a within-subject experiment. The independent t variable was sound: no sound, sonification. The second was graphics. We prepared three kinds of graphics for participants to copy (as shown in Fig. 10). In experiment one, 30 cm and 60 cm sizes graphics had lower deviation in 5 cent/cm and 10 cent/cm. We thought that 30 cm was too small in mid-air AR sketching. Therefore, the graphic size was defined as 60 cm and the pitch change rate was 10 cent/cm. Other settings were consistent with experiment one.

## B. Task and Procedure

Similar to experiment one, researchers first explained the method and purpose of the experiment to the participants and demonstrated the operation and process of the experiment to the participants. We set up practice tasks to help participants understand the relationship between spatial location and pitch. We arranged practice tasks before the formal experiment, and we avoided that the content of the practice experiment was the same as the formal experiment. Therefore, we took some basic graphic elements as the content of the practice experiment, such as circle, square, triangle and two sets of vertical lines with a size of 60 cm. Participants needed to stroke these graphics repeatedly, which was the same as the stroke task in experiment one. Participants had about 15 min to practice. When the participants were ready, the experiment began. Fig. 11 shows the screenshots at different experimental stages.

Participants were asked to try their best to control the graphic size at 60 cm. We did not provide any visual reference of 60 cm size, because this may affect the results. The order in which participants were exposed to sound conditions was no sound, sonification, sonification, and no sound. Within each sound condition, the order of the provided graphics was randomized. Overall, each participant copied 12 graphics. After all tasks, participants were asked to complete a NASA-LTX questionnaire. Each participant took about 20 min.

## C. Participants and Apparatus

Twelve students, half male, aged 21–27, were recruited. Eleven participants were right-handed and one participant were left-handed. Seven participants had more than one year of hand-painting experience, five participants had less than half a year of hand-painting experience. Ten participants had experience in using AR applications (not drawing application) before participating in the experiment.

Eight participants took part in both experiment one and experiment two. As there was a several-week break between the two experiments, participants in experiment two generally could not remember the sound feedback in experiment one. In addition, in experiment two, we did not fix the position of the graphic drawn by the participants in space, leading to different sound feedback for the same graphic. Therefore, experiment one should not have effect on the performance of participants in experiment two.

The experiment equipment was the same as experiment one. Each participant signed informed consent too.

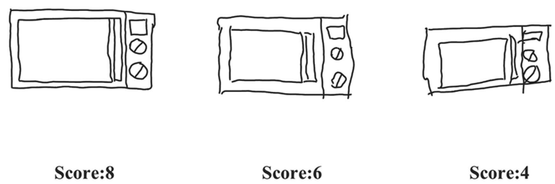



Fig. 12. Scoring reference standard in the first method.

### D. Measurements

Completion Time: The time for participants to complete one copy task.

Similarity: The degree of consistency between the drawn graphics and the reference graphics. It was used to evaluate the quality of the copied graphics. We used two different methods to calculate the similarity. In the first method, we invited 10 adjudicators (half male, aged from 18–30) to make a subjective evaluation of the consistency. The score ranges from 1 (very different) to 10 (very similar), and the adjudicators completed the evaluation through an online questionnaire. No participant's information and experimental conditions were provided in the questionnaire. We set a standard (see Fig. 12) to unify the evaluation criteria. In the second method, we chose Pdollarplus [30] to calculate the deviation between the copied graphic and the reference graphic. First, we used MATLAB to binarize the copied and reference graphics and resample them to 128 equally spaced points to construct candidate point clouds ($C$) and template point clouds ($T$). Resampling is common practice in order to uniformize input data for classifiers. Then, the Euclidean distance and turning angle were calculated to match the two point clouds (4). When the calculation result was the smallest, the points in the candidate point cloud ($C_i$) were matched with the points in the template point cloud ($T_j$). After all points were matched, the sum of the Euclidean distances between all matching points of $C$ and $T$ was calculated, which was the distance between the two point clouds. The source code of Pdollarplus is available online [39]

$$\|C_i - T_j\| = \sqrt{(C_{i.x} - T_{j.x})^2 + (C_{i.y} - T_{j.y})^2 + (C_{i.\theta} - T_{j.\theta})^2}. \quad (4)$$

Aspect Ratio Error: The difference in the aspect ratio between the reference graphic and the copied graphic. An accurate ratio is the basis for drawing complex graphics. We calculated the aspect ratio error to examine the effect of sonification on the graphic ratio in the copy task. The aspect ratio of reference graphic "microwave," "coffee cup," and "sunflower" is 2, 0.625, and 1.

Task Load Index Questionnaire: NASA-TLX was used to measure the effect of sonification on participants' workload based on six dimensions: mental demand, physical demand, temporal demand, subjective performance, effect, and frustration. The score ranges from 0 (*low level*) to 100 (*high level*). There were 20 levels in each dimension.

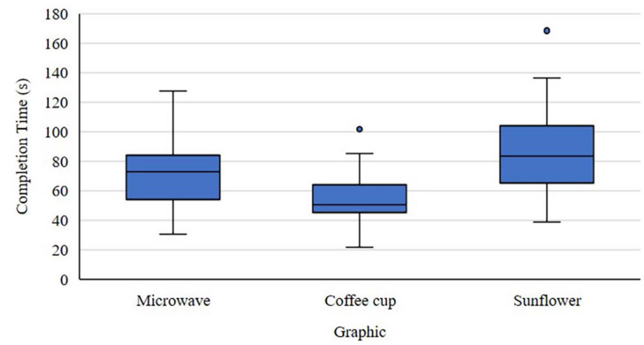



Fig. 13. Completion time of three reference graphics. Outliers based on the IQR rule.

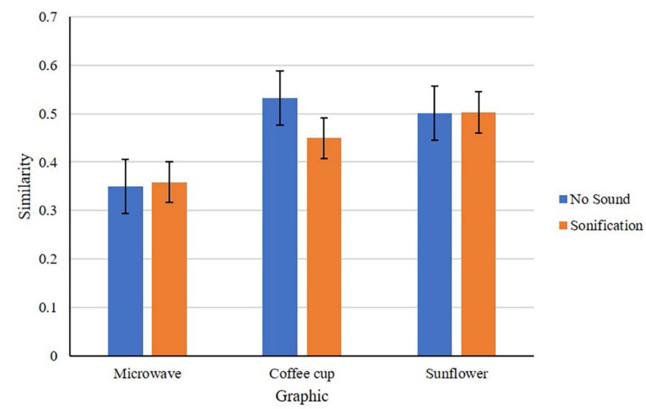



Fig. 14. Similarity of three graphics in tow sound conditions. During the calculation, the x and y of those points were normalized. Error bar was the 95% confidence interval.

### E. Results

Paired sample $t$-test was chosen to analyze data. The data that failed the normality test were analyzed by Wilcoxon signed rank test and Friedman test.

*1) Completion Time:* The sound variable had no significant effect on the completion time, $Z = -0.93$, $p = 0.35$. Completion time depended on graphic types, $\chi^2$ (2) $= 64.50$, $p < 0.01$ (see Fig. 13). There were significant differences between "microwave" (*Median* = 72.55 s), "coffee cup" (*Median* = 50.36 s) and "sunflower" (*Median* = 83.40 s).

*2) Similarity:* In method one, 10 adjudicators completed the questionnaire. Inter-rater reliability (Krippendorff's alpha coefficient) was 0.48. According to Landis *et al.* [23], the relative strength of agreement is "Moderate." However, the results showed that Sound variable had no effect on the similarity among the three kinds of graphics.

The results of method two showed that the sound variable had no significant effect on the similarity in "microwave" and "sunflower". Only "coffee cup", there was a marginally significant preference, $t$ (23) $= 1.96$, $p = 0.06$ (see Fig. 14). The similarity in sonification condition ($M = 0.45$) was lower than that in no sound condition ($M = 0.53$).

*2) Aspect Ratio Error:* The aspect ratio error of "microwave" was significantly lower in sonification condition ($M = 0.16$) than

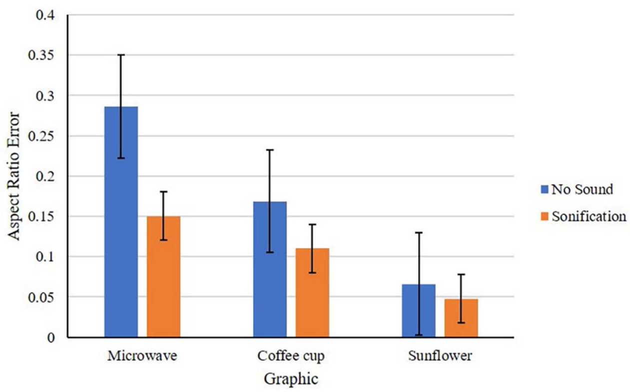



Fig. 15. Aspect ratio error of three graphics in different conditions. Error bar was the 95% confidence interval.

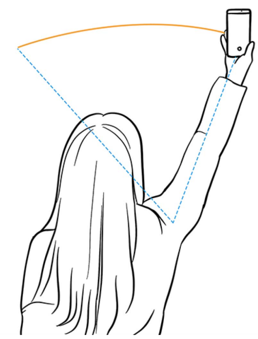


Fig. 16. When drawing large-scale graphics, the stroke are arcs because of shoulder joint rotation.

TABLE III
RESULTS OF THE NASA-TLX QUESTIONNAIRE

| Task Load Index Dimensions | Conditions | | | |
|---|---|---|---|---|
| | No Sound | | Sonification | |
| | Avg. | Std. | Avg. | Std. |
| Mental Demand | 65.83 | 13.29 | 53.75 | 14.48 |
| Physical Demand | 77.08 | 13.73 | 74.17 | 14.75 |
| Temporal Demand | 52.08 | 18.64 | 52.50 | 19.48 |
| Subjective Performance | 54.17 | 16.76 | 65.83 | 20.32 |
| Effort | 72.50 | 8.66 | 64.58 | 10.97 |
| Frustration | 57.92 | 20.72 | 51.25 | 21.01 |

in no sound condition ($M = 0.29$), $t\ (23) = 2.04$, $p < 0.05$. A significant effect was observed in "coffee cup". The sonification condition ($M = 0.11$) was significantly lower than the no sound condition ($M = 0.17$), $t\ (23) = 2.14$, $p < 0.05$. There was no significant difference for "sunflower" (see Fig. 15).

*4) Task Load Index:* The sound variable had significant effects on three dimensions: mental demand, subjective performance, and effect (see Table III). In sonification condition, the dimensions of mental demand ($Z = -2.55$, $p < 0.05$) and effect ($Z = -2.21$, $p < 0.05$) were significantly reduced, and the subjective performance ($Z = -2.14$, $p < 0.05$) was significantly improved.

## VI. Discussion

Through two user experiments, we have demonstrated that sonification can improve the accuracy of mid-air sketching with a mobile AR device. The results of experiment one showed that through repeated practice, the application of sonification can significantly reduce the graphic deviation caused by narrow FOV. The results of experiment two suggested that the participants were able to achieve a more accurate aspect ratio with real-time auditory feedback.

### A. Experiment One

In experiment one, the completion time was only affected by the graphic size. In the sonification condition, the completion time is neither increased nor decreased. It is only related to the length of the path. This result is the same as Gao *et al.* [28].

The deviation of the three graphic sizes all showed a downward trend. The deviation decreased by 28.5% at most in 30 cm size graphics. In 60 cm size graphics, it decreased by 26.8% at most. Although there is no significant reduction in 90 cm size graphics, the deviation came down gradually with the increase of pitch change rate. This is the same as Gao *et al.*'s conclusion [28] that sonification can indeed reduce the drawing deviation caused by the lack of visual information. In the experiment, we observed that the participants' arm was fully extended when drawing 90 cm size graphics. Because of the rotation of the shoulder joint, the stroke was not a straight line, but an arc (see Fig. 16). This is similar to Arora *et al.*'s [22] conclusion that drawing large-scale graphics in virtual space leads to a decrease in accuracy, which can be attributed in part to the fact that our arms follow natural arc motions. Participants tried to draw straighter lines by moving side to side or bending their knees, but this is difficult because it is hard to keep the arm stable during body movement and there will be a slight change in perspective with the movement. At present, only a few studies discussed the consequence of perspective change on drawing. Arora *et al.* [22] studied the effect of drawing plane orientations on drawing deviation, and the results showed that the deviation was lower in the horizontal and vertical conditions, but higher when the plane moved sideways. A similar conclusion was reached by Bea *et al.* [29], the perspective suitable for drawing is that the drawing plane has a larger projection area.

### B. Experiment Two

The results of experiment two showed that the more strokes, the longer the trajectory, and it takes more time. This is similar to the conclusion of experiment one.

In experiment one, participants were familiar with the auditory feedback through three-stroke tasks. In experiment two, participants only completed the practice of basic drawing elements, so they generally lacked the experience of the reference graphics. The drawn graphics in the sonification condition had a higher accurate aspect ratio. These two results showed that repeated practice is necessary to achieve better results. But without repeated practice, sonification can still provide additional information to help participants get a better aspect ratio.

In experiment two, we used two different methods for similarity evaluation. The subjective evaluation showed no significant difference in similarity, however, the Pdollarplus algorithm results showed marginally significance in "coffee cup". This may be due to the fact that the aspect ratio of the "coffee cup" is less than one, similar to the aspect ratio of most mobile phone screens. As a result, during the drawing process, participants were able to see more context about "coffee cup," resulting in better results. For "microwave" and "sunflower," holding the phone upright has less graphic context, so the results were not significant. NASA-TLX questionnaire showed that the subjective performance of the participants increased significantly, whereas the mental demand and effort decreased significantly in the sonification condition. This showed that in the opinion of the adjudicators, sonification cannot help the participants draw more accurately, but the participants' workload was reduced and their subjective feelings were better. This can effectively reduce the fatigue of designers in long-time drawing tasks.

### C. General Discussion

Aiming to alleviate the problem of narrow FOV of mobile AR devices, we apply sonification to AR drawing system design. We used sonification interaction to compensate for the graphic information through sound in the case of a lack of visual information. In our experiments, participants can perceive the spatial position of the device, the length and direction of strokes through pitch changes, so as to achieve a more accurate control effect. Participants obtained sound feedback corresponding to reference graphics through practice, and gradually established the graphic-sound correlation. For example, when drawing a horizontal or vertical line, the pitch of one sound remains stable and the pitch of the other sound gradually rises (or falls). When drawing a right angle, the pitch of one sound changes from constant to rising (or falling), and the pitch of the other sound changes from rising (or falling) to constant. The circle has a special correspondence, in which the change of one sound maintains a phase difference of 90° with the other [19]. It is not easy to understand or realize this tonal relationship, but participants can still determine a range for the circle through the pitch change, so as to avoid the excessive aspect ratio from becoming an ellipse.

We used two experiments to verify whether sonification can help to achieve better accuracy of mid-air sketching with mobile AR devices. Our first experiment showed that through repeated practice, the drawing deviation was reduced by 28.5% in 30 cm size graphics and 26.8% in 60 cm size graphics. The second experimental results showed that participants still obtained a better aspect ratio even if they lacked the practice process of reference graphics. At the same time, NASA-TLX questionnaire showed that under the sonification condition, the subjective performance of participants increased significantly, and the mental demand and efforts decreased significantly.

Through the above discussion, we believe that sonification has a positive effect on mid-air AR sketching, because it can effectively reduce the drawing deviation and the workload. If we want to achieve better drawing accuracy, the repeated practice is necessary. Participants could be familiar with the auditory feedback of relevant graphics and establish a correct graphic-sound model. Even without repeated practice, participants were still able to obtain a high aspect ratio.

In addition, some participants reported that the two timbres were similar, making it difficult to distinguish them in experiments. And continuous sound feedback in this study did not sound good, which made participants feel uncomfortable and affected their attention. This confirms the point of Ferguson and Brewster [25] and Walker [26], which revealed that the goal of sonification design is not only "intuitive and precise," but also "pleasure." An effective sonification design comes from the compromise of these three goals. Focusing only on the "intuitive and precise" may lead to the decrease in the stimulus-response compatibility (how natural a response feels based on a stimulus) [30], and may lead the user to give up the sound feedback completely. We use pitch-coordinate mapping because they have the same polarity, which is as simple as the metaphorical association between louder and more. However, the relationship between the strokes inside the graphics and the spatial relationship between graphics is much more complex. This may result in performance limitations when drawing complex graphics.

Based on our results in this study, we summarized the following implications that could be helpful for applying sonification to mid-air sketching with mobile AR devices.

1) Sonification does not increase the workload of users. The results showed that sonification did not affect the completion time, but reduced the mental demand and effort. Therefore, designers do not need to worry about increasing the workload of users when applying sonification.
2) The effect should be better when the graphic size less than 90 cm. The deviation did not decrease significantly when the graphic size was 90 cm, but decreased by 28.5% and 26.8% at the 60 cm and 30 cm, respectively. Thus, when applying sonification to mid-air AR sketching, the size of a single graph should be set within 90 cm.
3) Different pitch change rates are suitable for different graphic sizes. The results showed that the deviation of the 30 cm graphics was the lowest at the pitch change rate of 5 cent/cm, whereas the deviation of the 60 cm graphic was the lowest at the pitch change rate of 10 cent/cm. Therefore, different pitch change rates should be selected for different graphic sizes to achieve the best effect.
4) After repeated practice, users can improve their performance. Thus, users should be allowed to practice the reference graphics in order to establish the graphic-sound model.

## VII. Conclusion

Design in AR is not limited by physical plane, which can help designers create imaginative compositions. However, using mobile AR devices for mid-air sketching has the problem of narrow FOV, which leads to users' inability to see the context and drawing deviation. This article shows how to use sonification to alleviate the problem and help users draw accurate graphics. Our experimental results showed that sonification can significantly increase the drawing accuracy without extending the drawing time. In the condition of sonification, the drawn graphics had a more accurate aspect ratio. NASA-TLX questionnaire showed that the mental demand and effort were significantly reduced and the subjective performance was significantly increased. Our work provides a new solution to such problems and has good effects.

## VIII. Future Work

As the first study to use sonification to solve the problem of narrow FOV, we chose parameter mapping sonification, one of the most commonly used sonification techniques. However, as we mentioned in the discussion, this sonification design can only meet the goals of " intuitive" and "precise," but cannot meet the goal of "pleasure." The sound would affect the efficiency and accuracy of AR sketching. Moreover, this design can only indicate the position changes of the AR device. The internal structure of graphics and the spatial relationship between graphics cannot be displayed. Therefore, the future work should further explore the new sonification design suitable for graphics, and optimize the auditory experience.

We chose 2-D graphics as our research goal because 2-D graphics are easier to draw and widely used. However, in 3-D space, drawing 3-D graphics is a common task as well. Three-dimensional graphics are more expressive, but drawing 3-D graphics is challenging. Therefore, it will be a part of future work to expand our work and study the effect of sonification on drawing 3-D graphics in AR.

In this study, we compared the performance of drawing deviation under no sound and sonification conditions. However, this did not show that our work has advantages over other related work. Therefore, comparing sonification with other drawing techniques and comparing different sonification techniques will be part of the future work. In particular, we believe that sonification is not contradictory to other nonauditory drawing technologies. Therefore, integrating them into the AR drawing system with sonification may improve drawing efficacy.

## References


[1] K. C. Kwan and H. Fu, "Mobi3dsketch: 3D sketching in mobile AR," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–11.

[2] G. Krammer et al., "Sonification report: Status of the field and research agenda. Presented at ICAD," Mar. 2010. [Online]. Available: https://core.ac.uk/download/pdf/17237285.pdf

[3] G. Varni et al., "Interactive sonification of synchronisation of motoric behaviour in social active listening to music with mobile devices," *J. Multimodal User Interfaces*, vol. 5, no. 3-4, pp. 157–173, May 2012, doi: 10.1007/s12193-011-0079-z.

[4] *The Sonification Handbook*, 11/2011 ed. Berlin, Germany: Logos Publishing House, 2011, pp. 301–425.

[5] N. Schaffert, K. Mattes, and A. O. Effenberg, "The sound of rowing stroke cycles as acoustic feedback. Presented at ICAD," Jun. 2011. [Online]. Available: https://smartech.gatech.edu/handle/1853/51919

[6] D. Cesarini, T. Hermann, and B. Ungerechts, "An interactive sonification system for swimming evaluated by users. Presented at SoniHED," 2014. [Online]. Available: https://www.york.ac.uk/media/c2d2/media/sonihedconference/Cesarini_etal_SoniHED_2014.pdf

[7] J. W. Newbold et al., "Musically informed sonification for chronic pain rehabilitation: Facilitating progress & avoiding over-doing," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2016, pp. 5698–5703.

[8] J. Harris et al., "Sonic respiration, controlling respiration rate through auditory biofeedback," in *Proc. Extended Abstr. Hum. Factors Comput. Syst.*, Apr. 2014, pp. 2383–2388, doi: 10.1145/2559206.2581233.

[9] I. Hussain et al., "A comparative study of sonification methods to represent distance and forward-direction in pedestrian navigation," *Int. J. Hum.-Comput. Int.*, vol. 30, no. 9, pp. 740–751, Jul. 2014.

[10] I. Hussain et al., "Chinese-based spearcons: Improving pedestrian navigation performance in eyes-free environment," *Int. J. Hum.-Comput. Int.*, vol. 32, no. 6, pp. 460–469, Mar. 2016.

[11] W. Heuten, D. Wichmann, and S. Boll, "Interactive 3D sonification for the exploration of city maps," in *Proc. 4th Nordic Conf. Hum.-Comput. Interaction: Changing Roles*, 2006, pp. 155–164, doi: 10.1145/1182475.1182492.

[12] T. Smith et al., "Exploring gesture sonification to support reflective craft practice," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2015, pp. 67–76.

[13] Y. C. Han and P. Surve, "Eyes: Iris sonification and interactive biometric art," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–4.A.

[14] J. Cherston et al., "Sonification platform for interaction with real-time particle collision data from the atlas detector," in *Proc. CHI Conf. Extended Abstr. Hum. Factors Comput. Syst.*, 2016, pp. 1647–1653, doi: 10.1145/2851581.2892295.

[15] D. L. Mansur, M. M. Blattner, and K. I. Joy, "Sound graphs: A numerical data analysis method for the blind," *J. Med. Syst.*, vol. 9, no. 3, pp. 163–174, Jun. 1985, doi: 10.1007/BF00996201.

[16] J. Kim, Y. Lee, and I. Seo, "Math graphs for the visually impaired: Audio presentation of elements of mathematical graphs," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–6.

[17] S. Harada, H. Takagi, and C. Asalawa, "On the audio representation of radial direction," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2011, pp. 2779–2788.

[18] P. B. L. Meijer, "An experimental system for auditory image representations," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 2, pp. 112–121, Feb. 1992, doi: 10.1109/10.121642.

[19] D. Rigas and J. Alty, "The rising pitch metaphor: An empirical study," *Int. J. Hum.-Comput. Stud.*, vol. 62, no. 1, pp. 1–20, Jan. 2005, doi: 10.1016/j.ijhcs.2004.06.004.

[20] R. Arora et al., "Symbiosissketch: Combining 2D & 3D sketching for designing detailed 3D objects in situ," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2018, pp. 1–15.

[21] G. Dubus and R. Bresin, "A systematic review of mapping strategies for the sonification of physical quantities," *PLoS One*, vol. 8, no. 12, Dec. 2013, Art. no. e82491, doi: 10.1371/journal.pone.0082491.

[22] R. Arora et al., "Experimental evaluation of sketching on surfaces in VR," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2017, pp. 5643–5654.

[23] J. R. Landis and G. K. Gary, "The measurement of observer agreement for categorical data," *Biometrics*, vol. vol. 33, pp. 159–174, Mar. 1977.

[24] G. Parseihian, C. Gondre, M. Aramaki, S. Ystad, and R. Kronland-Martinet, "Comparison and evaluation of sonification strategies for guidance tasks," *IEEE Trans. Multimedia*, vol. 18, no. 4, pp. 674–686, Apr. 2016.

[25] J. Ferguson and S. Brewster, "Modulating personal audio to convey information," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–6.

[26] B. N. Walker, "Magnitude estimation of conceptual data dimensions for use in sonification," *J. Exp. Psychol.*, vol. 8, no. 4, Jan. 2003, Art. no. 211.

[27] B. N. Walker and A. Ehrenstein, "Pitch and pitch change interact in auditory displays," *J. Exp. Psychol.*, vol. 6, no. 1, Apr. 2000, Art. no. 15.

[28] B. Gao, H. Kim, H. Lee, J. Lee, and J. -I. Kim, "Effects of continuous auditory feedback on drawing trajectory-based finger gestures," *IEEE Trans. Hum.-Mach. Syst.*, vol. 48, no. 6, pp. 658–669, Dec. 2018, doi: 10.1109/THMS.2018.2850329.

[29] S. H. Bae, R. Balakrishnan, and K. Singh, "ILoveSketch: As-natural-as-possible sketching system for creating 3D curve models," in *Proc. 21st Annu. ACM Symp. User Interface Softw. Technol.*, 2008, pp. 151–160.

[30] R. D. Vatavu, "Improving gesture recognition accuracy on touch screens for users with low vision," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2017, pp. 4667–4679.

[31] B. Nuernberger et al., "Snaptoreality: Aligning augmented reality to the real world," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2016, pp. 1233–1284.

[32] D. Gasques et al., "What you sketch is what you get: Quick and easy augmented reality prototyping with pintar," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–6.

[33] P. Wacker et al., "Arpen: Mid-air object manipulation techniques for a bimanual ar system with pen & smartphone," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–12.

[34] V. Nair and B. A. Smith, "Toward self-directed navigation for people with visual impairments," in *Proc. Adjunct Publication 33rd Annu. ACM Symp. User Interface Softw. Technol.*, 2020, pp. 139–141.

[35] A. Botzer, N. Shvalb, and B. B. M. She, "Using sound feedback to help blind people navigate," in *Proc. 36th Eur. Conf. Cogn. Ergonom.*, 2018, pp. 1–3.

[36] J. J. LaViola and D. F. Keefe, "3D Spatial interaction: Applications for art, design, and science," in *Proc. SIGGRAPH*, 2011, pp. 1–75.

[37] S. Kratz and M. Rohs, "The $3 recognizer: Simple 3D gesture recognition on mobile devices," in *Proc. Int. Conf. Intell. User Interfaces*, 2010, pp. 419–420.

[38] B. Jens, B. Denis, and W. Gerhard, "Comparing computer-based drawing methods for blind people with real-time tactile feedback," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2018, pp. 1–13.

[39] $P+ Point-Cloud Recognizer. [Online]. Available: http://depts.washington.edu/acelab/proj/dollar/pdollarplus.html

[40] GarageBand. [Online]. Available: https://www.apple.com.cn/ios/garageband/

[41] ARCore. [Online]. Available: https://developers.google.cn/ar/


**Haonan Xu** is currently working toward the Postgraduate degree with Beijing University of Posts and Telecommunications, Beijing, China.

He is currently studying information and interaction design with the School of Digital Media and Design Arts, Beijing University of Posts and Telecommunications. His research interest is human-computer interaction.

**Fei Lyu** received the Ph.D. degree from the University of Chinese Academy of Sciences, Beijing, China, in 2013.

She is an Associate Professor with the School of Digital Media and Design Arts, Beijing University of Posts and Telecommunications, Beijing, China. She has authored or coauthored more than 20 research papers including top-tier HCI conference papers (e.g., ACM CHI, ACM CSCW) and journal papers such as IJHCS. Her research area include human-computer interaction, with special interests in natural user interface and child-computer interaction.

**Jin Huang** received the Ph.D. degree from the Department of Computer Science, University of Chinese Academy of Sciences, Beijing, China, in 2018.

He is an Associate Research Fellow with the Intelligence Engineering Laboratory, the Institute of Software, Chinese Academy of Sciences, Beijing, China. He was an Academic Visitor with the Department of Engineering, University of Cambridge from Jan. 2020 to Apr. 2021. He has authored or coauthored more than 40 research papers including top-tier HCI conference/journal papers such as *ACM CHI, ACM UIST, IJHCS*. His research interests include the general area of human-computer interaction, with a primary focus on understanding and modeling human behaviors with praxeology, psychology and artificial intelligence.

Dr. Huang has served as a program committee member for ACM CHI 2020-2022. He is a Reviewer of top-tier HCI conference and journal, such as *ACM CHI, ACM UIST, IJHCS, Ubicomp* and IEEE VR.

**Huawei Tu** received the Ph.D. degree from Kochi University of Technology, Kochi, Japan, in 2012.

He is a Lecturer with La Trobe University, Bundoora, Australia. He was a Postdoc Researcher with Ritsumeikan University of Japan from 2012 to 2013, and with Swansea University of the U.K. from 2013 to 2014. Before joining La Trobe University, he was an Associate Professor with Nanjing University of Aeronautics and Astronautics in China from 2014 to 2019. His research interests include human-computer interaction (HCI) and virtual reality (VR), with special interests in multimodal interaction, user interface design, and user experience design. He has authored or coauthored more than 50 research papers including top-tier HCI journal papers (e.g., *ACM TOCHI*, IEEE TVCG) and conference papers such as *ACM CHI* and IEEE VR.

Dr. Tu has served as a Program Committee Member for ACM ISMAR, IEEE ICVR 2021, IWAMSM 2020, Chinese CHI 2013-2021, ACM IUI 2019, IxAP & IDHF 2018-2014, APCHI 2012-2014. He is the Secretary-General and a Committee Member of International Chinese Association of Computer Human Interaction. He was a visiting Researcher with Kochi University of Technology in Japan from July 2017 to March 2020. He was a Guest Editor of an IWC special issue (2016) and CCF Transactions on Pervasive Computing and Interaction (2020).