

# Interactive Image Exploration for Visually Impaired Readers using Audio-augmented Touch Gestures

Rynhardt Kruger<sup>\*†</sup>, Febe de Wet<sup>\*</sup> and Thomas Niesler<sup>\*</sup>

<sup>\*</sup>*Department of Electrical and Electronic Engineering*

*Stellenbosch University, South Africa*

rkruger,fdw,trn@sun.ac.za

<sup>†</sup>*Digital Audio-Visual Technologies, Networked Systems and Applications*

*Next Generation Enterprises and Institutions Cluster, CSIR, South Africa*

rkruger@csir.co.za

**Abstract**—Technologies such as text to speech and hardware braille displays provide alternative representations of electronic text. As a result, electronic documents provide increased accessibility to print-disabled users. However, non-textual graphical information in electronic documents remain largely inaccessible to the blind population. In this study, we explore audio-visual sensory substitution as a means of rendering graphical information present in electronic documents to blind users. To achieve this, we have extended the audio rendering approach used by the well-established vOICe algorithm to allow interactive and localised exploration of an image by means of gestures and the touch screen of a standard commercially-available tablet. The effectiveness of our approach was evaluated in a set of user trials that required six sighted and six blind subjects to identify elements of scenes consisting of a number of geometrical shapes and emoticons. Our results show that both groups of subjects were more successful at identifying shapes using the interactive algorithm than with the baseline vOICe algorithm to a highly statistically significant degree. Furthermore, the results indicate that this improvement is greatest for the most complex scenes. We conclude that, by introducing an interactive touch interface, the vOICe algorithm can be successfully extended to allow interactive exploration and interpretation of diagrams, thereby improving accessibility to material such as scientific publications.

**Index Terms**—Sensory substitution, vOICe algorithm, blindness, visual impairment, accessibility, diagrams.

## I. INTRODUCTION

Since the widespread adoption of electronic document formats, such as the Portable Document Format (PDF), printed material has become much more accessible to blind and visually-impaired users. This is because the text in such electronic documents can be read by screen readers incorporating text-to-speech or braille functionality [1], [2]. However, non-textual graphical content, such as diagrams, graphs and equations, remain inaccessible, since their structure is usually not encoded into the electronic document. Although standards exist for adding descriptions to graphical material (referred to as alt text), these are often not adhered to. Furthermore, textual descriptions may not be sufficient when a graphical representation is integral to the document, for example a geographical map or the specific shape of a signal. For this reason much technical writing, including contemporary published scientific research, remains largely inaccessible to the blind and visually impaired community.

Sensory substitution is the use of one sense to interpret information normally received by another. An example is the use of vibration to convey auditory information, which can be understood by people who are deaf or hard of hearing [3]. Sensory substitution is also used to convey sensory perceptions that are remote from the user, such as when performing robotic surgery [4].

We explore the possibility of using sensory substitution to allow blind or visually-impaired users to access graphical content encountered in an electronic document. Our approach is to render the graphical content as audio. To achieve this, we extend the well-established vOICe algorithm, which renders an image as a sequence of tone chords. The standard implementation of the vOICe algorithm was designed to render an image in its entirety. However, more targeted exploration is necessary for understanding complex images like scientific diagrams and graphs. To this end we incorporate gestures and a touch screen to allow flexible and interactive exploration of the rendered scene. This allows an image that is too complex to be interpreted at once to be systematically explored.

## II. BACKGROUND

Blind users are able to access electronic documents using software screen readers. Current accessibility standards require that graphical information be described in text, which can be read by screen readers or displayed in braille. An example is alt text that can be added to image elements on the world wide web, as described in the Web Content Accessibility Guidelines [5]. These descriptions are traditionally added manually, although recently pattern approaches have been used to generate some descriptions automatically [6], [7]. Textual descriptions of graphical content can also be deduced from the accessibility APIs of some software packages, such as a description of charts in Microsoft Excel [8]. However, a textual description on its own is often insufficient for describing graphical content where the representation is integral to understanding the content. An example is a depiction in which spatial relations are important, such as a geographical map or a floor plan. Graphical depictions are also often used to visualise large amounts of data in a succinct way, in which

case a detailed description of the diagram would negate the advantage of the representation.

A number of systems have been proposed to allow users to explore diagrams by separating the content into logical components. The Technical Drawings Understanding for the Blind (TeDUB) system allows users to explore circuit diagrams, Unified Modeling Language (UML) models, and architectural models by navigating through a tree structure containing a hierarchical view of the diagram's structure [9]. An extension to Scalable Vector Graphics (SVG) vector-based image formats, in which logical tags are added to diagram components, is described in [10]. This system uses text to speech to convey a component's general overview, type and shape, and haptic feedback to convey a component's location within the image.

Component-based exploration systems like the ones described above require an image to be interpreted and tagged before it can be explored. This interpretation can be achieved automatically, as is the case with the TeDUB system, or by manual tagging. However, automatic tagging is usually domain specific, and requires an extension to the program for each new diagram element, or class of diagrams. The TDUB system for example, can only interpret circuit diagrams, UML diagrams, and architectural plans. In order for blind people to access diagrams without manual preparation or domain-specific automatic tagging, a method providing direct access to the geometric composition of the diagram is required.

Specialised hardware devices have been developed to render graphical information in a form accessible to blind users. For example, the Iveo system by View Plus embosses SVG graphics onto Braille paper. When the resulting tactile diagram is placed on top of a touch pad, it offers a combined speech and tactile rendering of the diagram with the aid of specialised software [11] [12]. The Tactisplay Table device by Tactisplay Corp also produces a tactile rendering of a rasterised image by means of a specially-designed multiline electronic braille display [13].

Limitations of hardware-based approaches like those described above include that they are costly and are not easily portable. In this study, we focus specifically on the use of audio to convey graphical information that can not be represented by conventional screen readers. Audio has the advantage of not requiring specialised hardware, and thus being much more cost effective and accessible for blind users. Specifically, our algorithm is implemented on a consumer mobile tablet device.

The literature describes a few attempts to utilise a touch screen for audio-visual sensory substitution. Klatzky et al compared three approaches to touch-screen assisted diagram reading, namely vibration output, sound output, and a combination of the two [14]. In the first case, vibration is used to signal when a line is touched by the user's finger on the touch-screen. A specific vibration pattern is used to indicate the point where two lines meet. When using sound output, a sound is played when a line is touched, with stereo panning indicating the horizontal position of the finger on the screen and pitch the vertical position. A different sound is used to

denote an intersection of lines.

The study finds the main limitation of vibration output to be that the source of the stimulation is not well coordinated with the location being explored. The vibration is produced at a central point not spatially linked to the touch location, and thus cannot convey location or direction. This leads to the loss of contact with traced lines and other elements, as the user does not have advance warning of an approaching turning point or edge.

Klatzky *et al* also noted an increase in effectiveness when the focus, that is, the part of the image being rendered, is controllable by the user. One motivation for the research we describe in the following is the belief that more effective exploration of a displayed image can be achieved by allowing a greater degree of user interaction.

### III. THE VOICE

One of the first implementations of audio-visual sensory substitution was the Optophone, as described by Fournier D'Albe [15]. This electronic device divided the light reflected from a vertical slice of the input image into discrete segments. Moving from bottom to top, each segment was mapped to an increasingly higher frequency. This range of frequencies was then sonified with the help of a selenium photosensor. In this way a chord was emitted for each vertical slice, where bright parts are sonified at a frequency corresponding to the vertical position while dark parts are silent. Users were expected to read letters by listening for the missing frequencies.

Almost 80 years later, a method similar to that employed by the Optophone was encoded as a computer algorithm by Meijer [16]. This led to the software implementation known as the vOICE. Like the optophone, the vOICE produces a chord of frequencies for each column of the input image. Therefore, each frequency in the produced chord corresponds to a specific pixel in the current column, where the vertical height of the pixel determines the frequency (higher positions correspond to higher frequencies), and the brightness corresponds to loudness. By scanning the image from left to right, the chords are played in sequence, with a click denoting the end of the image. In later versions of the vOICE, the chords were played from left to right in the stereo field, such that the left most column of the image produces a sound most to the left and the right most column a sound most to the right [17].

### IV. PROPOSED ALGORITHM

We extend the vOICE, as described in the previous section, by taking particular advantage of a touch screen to allow interactive image exploration. We have considered specifically the challenges of visualising line drawings, as may be present in published technical material. When using the vOICE, the user hears a rendition of the entire image, and it is not possible to directly control what is rendered. This makes it difficult to explore details of the image, such as the connections in a flow chart or the evolution of a graph. While more recent versions of the vOICE do allow the user to define the area being rendered by indicating the upper-left and lower-right

corners on the touch screen, this feature was designed to focus the rendition of an everyday scene and is not appropriate for document reading. In particular, it is ill-suited to the tracing of lines or the finding of edges, as the scanning process results in a slow exploration.

It should be borne in mind that the vOICe was intended as a continuous vision substitution algorithm, for use when exploring the physical environment with a live camera. In such a situation, large details are important, and can be quickly identified when the camera view is scanned from left to right. In addition, the changing camera view as the user moves or alters their body posture serves as a natural way of exploring the environment. In contrast, when reading a diagram, the view is not expected to change. Instead, the user is expected to carefully explore various areas of the image interactively.

We extend the algorithm by allowing it to be controlled using two simple gestures. This allows the user to interactively explore details of the image. When one finger is placed on the screen, the system selects a short vertical segment of the image located directly under the user’s finger. This segment is then rendered as a chord using the vOICe algorithm. This allows precise local exploration of the image. When two fingers are placed on the screen, the segment along the line connecting these two locations is sonified as a chord. The upper and lower frequencies are determined by the vertical locations of the upper and lower finger respectively. Using this gesture, the user can control both the length of the segment (by moving the fingers closer together or further apart) and the orientation of the segment (by rotating the fingers). This allows line segments with arbitrary orientation to be located because, in contrast to the vOICe which uses fixed start and end points for the scanning path, these can now be arbitrarily and interactively defined. The algorithm is implemented on a conventional tablet computer, making it cost effective.

## V. EXPERIMENTAL EVALUATION

To evaluate the interactive algorithm, six blind and six sighted subjects were recruited and asked to identify a sequence of images presented to them as audio on a tablet computer. These images ranged from single simple geometric shapes to more complex compositions. Both the vOICe and the interactive algorithm were used to render the images. The objective was to determine whether the additional functionality offered by the interactive algorithm is in fact beneficial to the user. It was decided not to blindfold the sighted subjects during the test, since neither algorithm renders to the screen. The screen is used only as an input device.

### A. Test images

The test comprises five stages. The first three concern the identification of simple shapes (Figure 1) while the last two involve the identification of emoticons (Figure 2). The simple shapes in question are squares, circles, ovals (horizontal and vertical), rectangles (horizontal and vertical), and triangles in four orientations. The emoticons are a smiley face, a sad face, a skew-mouthed face, and a face with a winking right eye.

The five stages of the experimental evaluation are presented in Table I.

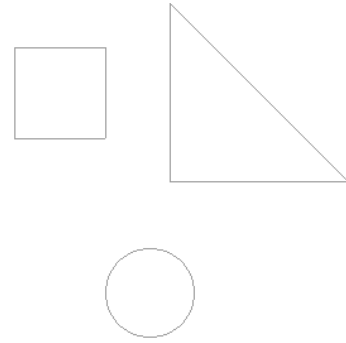


Fig. 1. The three categories of shapes used in Stages 1, 2 and 3: Triangle and square at the top, circle at the bottom.

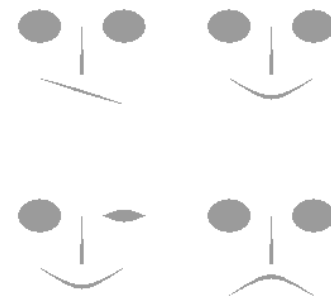


Fig. 2. The four emoticons rendered in Stages 4 and 5: Skew mouth and smiley at the top, sad face and winking face at the bottom.

Stage	Description	vOICe	Interactive	Total
1	Two simple shapes are rendered	2	2	4
2	Three simple shapes are rendered	3	3	6
3	Four simple shapes are rendered	3	3	6
4	A single emoticon is rendered	2	2	4
5	Four emoticons are rendered	5	5	10
All		15	15	30

TABLE I  
THE STAGES OF THE EXPERIMENTAL EVALUATION, SHOWING THE NUMBER OF TEST IMAGES PRESENTED USING THE vOICe AND INTERACTIVE ALGORITHMS RESPECTIVELY.

For each stage of testing, a number of images are synthesised. Each test image is paired with either the vOICe or the gesture-enabled algorithm, resulting in a sequence of image-algorithm pairs. The same image was never rendered by both algorithms to avoid possible sequential bias in the test. Users are required to listen to the image as rendered by the algorithm, after which a menu containing a list of shapes is presented. Users must then identify all shapes that they recognise in the image using this menu.

A shape can only be selected once, even if it appears more than once in the image. This decision was taken as a

compromise between the duration of the testing procedure, and the number of data points that can be collected. In preliminary tests, where users were required to provide more detailed descriptions of the rendered images, the duration of the test procedure proved too long for most test candidates. In its current form, the test procedure takes most subjects about an hour and a half to complete.

The shapes for Stages 1 through 3 were generated algorithmically and written to the rendering buffer. These shapes were outlined but not filled. The emoticons for Stages 4 and 5 were pre-rendered using Tikz, and consisted of filled circles for the eyes, a filled triangle for the nose, and a narrow filled ellipse for the mouth. The mouth was either curved or angled to produce the possible mouth shapes, while one eye was narrowed horizontally to produce the winking face.

### B. Test procedure

The test was managed by software that renders the test images using one of the two described algorithms. After rendering each image, the system presents a menu containing the names of graphical objects that might be present in the image. Using this menu, the test subject must identify the components present in the image. Each component can be selected only once, even if the image contains several instances of this component. This was done to keep the testing procedure simple. Initial informal experimentation with more detailed test responses resulted in a procedure that was impractically long.

The test management software produces a log of each test subject's session indicating the selections made, the correct answer, as well as additional data such as the time elapsed since the shape was rendered, and in the case of the interactive algorithm, the gestures used to explore the image.

To perform the test, the user is presented with a tablet computer, a pair of headphones, and a wireless keyboard. The software that manages the test runs on the tablet computer. All messages are presented as synthesised speech.

The test process consists of the following steps.

- 1) The program introduces itself to the test subject.
- 2) The test subject is informed of the stage of testing (for example, identify shapes, identify emoticons, etc).
- 3) From a set of predetermined possible image and algorithm pairs, an image is selected and rendered with one of the two algorithms. In the case of the interactive algorithm, the subject is invited to explore the image in more detail by using the touch screen. In the case of the vOICe, the scene is rendered.
- 4) A menu listing the possible components that may be present in the image is presented to the test subject. The menu also contains an option to listen to the image again, and to continue to the next question. The subject can move through the menu using the up and down cursor keys. The subject was required to identify the components present in the rendered image, but not their individual locations. This was done to reduce the

complexity of taking the test. Earlier, informal experimentation requiring more detailed feedback was found to lead to an excessively long test and fatigue among the test subjects.

- 5) The algorithm, along with the correct answer, the subject's answer, the number of times the image was explored, and the duration of each attempt is recorded to a log file.
- 6) Items 2 through 5 are repeated for each algorithm, stage of testing and configuration of test images. Stages are selected in sequence, and algorithms and images are selected from a randomised set of image and algorithm pairs.

All subjects were given between 60 and 90 minutes of individual training before taking the test. During this training, subjects were required to practice recognising lines and simple shapes from the corresponding sound patterns produced by the vOICe. Subjects were also taught to use the gestures offered by the interactive algorithm. Finally, test subjects were familiarised with the shapes that occur in the test, as well as the emoticons. This was important since most blind participants had never encountered these emoticons before. The test was started when a user declared him/herself confident in using the two algorithms. All test subjects are shown the same test images and each test image is always rendered using the same algorithm.

### C. Test subjects

Evaluations were performed with 12 test subjects (six male and six female) between the ages of 20 and 45. Hence each of the images described in Table I was considered by 12 subjects, leading to a total of 360 responses, 180 for the vOICe and 180 for the interactive algorithm respectively. Half of the subjects were legally blind while the remainder were sighted.

One of the blind subjects was from a scientific background, while another was employed as an assistive technology specialist. The remaining four had no scientific or engineering background. In contrast, all six of the sighted subjects were from a scientific background.

## VI. RESULTS

The blind subjects reported that they found the menu system easy to operate and use, as it was similar to TTS-based menus used by other assistive interfaces. All subjects, both sighted and blind, understood within a short time how the vOICe based image-to-sound mapping worked. Most subjects were also able to use the interactive algorithm within a few minutes of introduction, although confident use took some time.

Table II lists the results for each stage as attained by blind subjects, sighted subjects, and both groups combined. The table shows that, overall, the vOICe algorithm resulted in 13% correct responses, while the interactive algorithm resulted in 48% correct responses.

Figure 3 summarises the percentage correct responses per stage in the test. Note that the analysis presented in this figure is based on least squares means and hence deviates slightly

Stage	Blind subjects (6)		Sighted subjects (6)		All subjects (12)	
	vOICe	Interactive	vOICe	Interactive	vOICe	Interactive
1	0.0	16.7	0.0	33.3	0.0	25.0
2	5.6	22.2	5.6	44.4	5.6	33.3
3	22.2	16.7	0.0	38.9	11.1	27.8
4	66.7	83.3	50.0	91.7	58.3	87.5
5	10.0	53.3	0.0	70.0	5.0	61.7
Total	17.8	38.9	7.8	56.7	12.8	47.8
Average time (s)	62.0	159.4	66.6	141.8	64.3	150.6

TABLE II

PERCENTAGE CORRECT IDENTIFICATIONS AND AVERAGE RESPONSE TIMES PER TESTING STAGE AND ALGORITHM TYPE (VOICE AND INTERACTIVE).

from the results in Table II. The figure clearly shows that the interactive algorithm exhibits higher accuracy during all five stages of testing, but that this improvement is particularly large in Stage 5 (the identification of multiple emoticons). This is an encouraging result, since the composition of four emoticons in a single image can be considered to be the most complex among the images in the test.

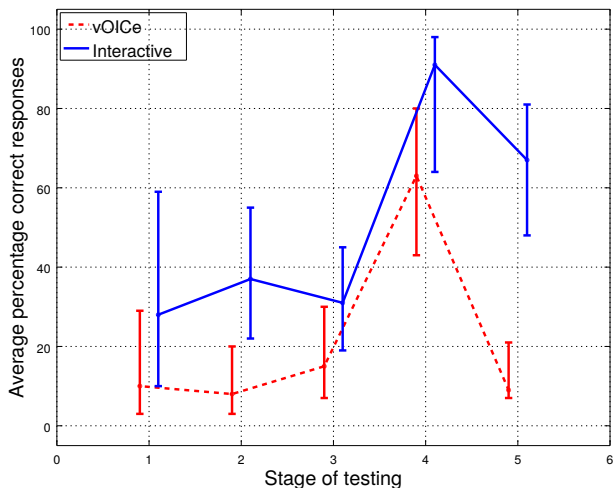


Fig. 3. Percentage of correct responses for each algorithm, for each stage of the test.

In our testing procedure, an answer was considered correct when the user correctly identified all components in the rendered image. If the user identified a component that was not present, or failed to recognise any of the components correctly, the answer was regarded as incorrect. This means that the table above represents only answers for which an image was perfectly described. We decided on this approach in an attempt to reduce the cognitive load on the subjects, since the duration of the test was already more than an hour in most cases. Due to this test structure, it was however not possible to unambiguously determine which shapes were most often confused.

## VII. DISCUSSION

As shown by the table in the previous section, both the blind and sighted groups performed better when interpreting the rendition of graphical information using the interactive algorithm, compared with the rendition produced by the vOICe. This was confirmed significant ( $p < 0.01$ ) using both a Chi-squared

and a Fisher exact test with Rao-Scott adjustment to account for each subject contributing more than one response. This significance holds both overall as well as when considering blind and sighted subjects separately and suggests that the touch screen gestures provide a clear additional benefit when reading complex shapes.

During training, a number of blind subjects reported to experience occasional difficulty in interpreting the sound produced by the interactive algorithm. It was observed that these subjects were inadvertently rotating their fingers into a slightly diagonal position, while believing them to be vertically aligned. Since the interactive algorithm renders the pixels on the line between the two fingers continuously as the fingers are moved across the screen, such diagonal placement leads to a vertical line being rendered as a rising or falling pitch. Reminding subjects to pay careful attention to the alignment of their fingers usually improved the situation. Nevertheless, this aspect of tactile exploration on a touch screen by blind users merits further consideration.

Table II also reveals that the blind subjects were better than the sighted subjects at interpreting the renditions produced by the vOICe, in which exploration gestures were not allowed. Although both groups did better when using the interactive algorithm than when using the vOICe, the sighted users achieved higher scores with the interactive algorithm overall.

There are two possible explanations for this. First, the sighted subjects might have had an advantage since they could see their fingers move on the touch screen and thereby avoid, for example, the inadvertent rotation described above. Second, the sighted subjects might simply be more practised in interpreting two-dimensional visual information. For example, during the tests it was noticed that a number of the blind users had to be familiarised with the concept and shape of an emoticon since they had not encountered it before. This seems to support the second hypothesis. However, if the only advantage enjoyed by the sighted subjects was their much greater familiarity with the interpretation of two-dimensional shapes, one would expect them also to do better when recognising renditions produced by the vOICe. Therefore, we suspect that the sighted subjects benefited also from their ability to follow their fingers when using the interactive algorithm. Sighted users were able to more effectively visualise shapes by combining accurate knowledge of their finger positions with their familiarity of the rendered shapes. In fact, two sighted users did report that they were able to visualise the shapes as

if they had appeared in front of them. This was not reported by any of the blind subjects.

It should also be borne in mind that most of the sighted subjects were from a scientific background (five had completed an undergraduate degree in science, while the sixth was enrolled in an engineering degree). In contrast, the blind subjects were mostly from a humanities background, with the exception of one who has a masters degree in computer science. Although not shown in the table of results, the blind subject from the computer science background attained a score of 10 for the interactive algorithm, which was similar to the scores achieved by the sighted group. It is therefore also possible that the scientific background of the sighted subjects allowed them to infer the shapes more easily. However, this scientific knowledge did not seem to aid them in interpreting the renditions produced by the vOICE.

Feedback provided by the subjects indicated that the users had most difficulty differentiating between the circle and ovals, the square and rectangles, as well as the smiley and winking face. The confusion between the smiley and winking face is understandable, since the two shapes are very similar, differing only in the right eye, which is narrower on the winking face. The confusion between the circle and ovals, as well as between the square and rectangles, may be explained by the fact that the overall soundscapes produced in both cases are similar to each other, differing only in the width of the frequency spectrum, as well as the duration in time. For example, both a square and rectangle consists of a click, followed by two simultaneous tones (one higher than the other), followed by another click. The difference lies in the duration between the first and final click, as well as the difference between the frequencies of the two tones.

## VIII. CONCLUSIONS AND FUTURE WORK

In this study we have considered audio-visual sensory substitution as an aid for rendering diagrams to blind users. We have extended the audio rendering approach used by the well-established vOICE algorithm to allow interactive and localised exploration of an image by means of gestures and a touch screen. The effectiveness of our approach was evaluated by means of a set of user trials with six sighted and six blind subjects to identify the elements of scenes consisting of a number of geometrical shapes and emoticons. We found that both groups of subjects were more successful at identifying shapes using the interactive algorithm than with the baseline vOICE algorithm.

An aspect that will receive attention in our ongoing work is the observed occasional tendency of blind users to misjudge the vertical alignment of their fingers when applying the gestures. Specifically, if the user believes two fingers to be vertically aligned while they are not, the audio scene can be misinterpreted. Another aspect that we are devoting attention to is the rendering as speech of any textual elements embedded in the diagram, such as the labels on a graph. Such elements can be identified in a PDF when the diagram is rendered as

vector graphics, and would further aid the interpretation of such figures in scientific material.

## IX. ACKNOWLEDGMENTS

The first author was supported financially by the South African Council for Scientific and Industrial Research (CSIR) to pursue his PhD.

## REFERENCES

- [1] G. Evans and P. Blenkhorn, "Screen readers and screen magnifiers," in *Assistive Technology for Visually Impaired and Blind People*, M. A. Hersh and M. A. Johnson, Eds. London: Springer London, 2008, pp. 449–495.
- [2] Freedom Scientific, "Jaws screen reader documentation," <https://support.freedomscientific.com/Products/Blindness/JAWSdocumentation>, Last accessed: 6 July 2020.
- [3] S. Levänen, V. Jousmäki, and R. Hari, "Vibration-induced auditory-cortex activation in a congenitally deaf adult," *Current Biology*, vol. 8, no. 15, pp. 869–872, 1998.
- [4] A. I. Aviles-Rivero, S. M. Alsaleh, J. Philbeck, S. P. Raventos, N. Younes, J. K. Hahn, and A. Casals, "Sensory substitution for force feedback recovery: A perception experimental study," *ACM Transactions on Applied Perception*, vol. 15, no. 3, pp. 16:1–16:19, Apr. 2018. [Online]. Available: <http://doi.acm.org/10.1145/3176642>
- [5] World Wide Web Consortium, "Web content accessibility guidelines (WCAG) 2.0," <https://www.w3.org/TR/WCAG20/>, 2008, Last accessed: 10 July 2020.
- [6] D. Guinness, E. Cutrell, and M. R. Morris, "Caption crawler: Enabling reusable alternative text descriptions using reverse image search," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–11.
- [7] S. Wu, J. Wieland, O. Farivar, and J. Schiller, "Automatic alt-text: Computer-generated image descriptions for blind users on a social network service," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 2017, pp. 1180–1192.
- [8] S. Gupta, M. V. Belani, D. Kaushal, and M. Balakrishnan, "Microsoft Excel charts accessibility: An affordable and effective solution," in *Proceedings of the 3rd International Workshop on Digitization and E-Inclusion in Mathematics and Science (DEIMS)*, 2016.
- [9] M. Horstmann, M. Lorenz, A. Watkowski, G. Ioannidis, O. Herzog, A. King, D. G. Evans, C. Hagen, C. Schlieder, A.-M. Burn, N. King, H. Petrie, S. Dijkstra, and D. Crombie, "Automated interpretation and accessible presentation of technical diagrams for blind people," *New Review of Hypermedia and Multimedia*, vol. 10, no. 2, pp. 141–163, 2004.
- [10] P. Roth, T. Pun, and J. Kronegg, "Rendering digital images accessible for blind computer users," in *HCI International 2003, 10th International Conference on Human-Computer Interaction*, 2003.
- [11] J. A. Gardner, V. Bulatov, and H. Stowell, "The ViewPlus IVEO technology for universally usable graphical information," in *Proceedings of the 2005 CSUN International Conference on Technology and People with Disabilities*, 2005.
- [12] ViewPlus, "Iveo 3 hands-on learning system," <https://viewplus.com/product/iveo-3-hands-on-learning-system/>, 2005, Last accessed: 1 November 2019.
- [13] Tactisplay Corp, "Tactisplay Table," <http://www.tactisplay.com/product/tactisplay-table>, 2015, Last accessed: 1 November 2019.
- [14] R. L. Klatzky, N. A. Giudice, C. R. Bennett, and J. M. Loomis, "Touch-screen technology for the dynamic display of 2D spatial information without vision: promise and progress," *Multisensory Research*, vol. 27 5-6, pp. 359–78, 2014.
- [15] E. E. Fournier D'Albe, "On a Type-Reading Optophone," *Proceedings of the Royal Society of London Series A*, vol. 90, pp. 373–375, Jul. 1914.
- [16] P. B. L. Meijer, "An experimental system for auditory image representations," *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 2, pp. 112–121, 1992.
- [17] —, "Seeing with sound for the blind: Is it vision?" in *Tucson Conference on Consciousness, April 8-12, 2002*, p. 83.