AMSE JOURNALS-AMSE IFRATH Publication –2016- Modelling C; Vol. 77; N° 2; pp 155-168
Submitted Sept. 2016; Accepted Oct. 31, 2016

Non-Visual Selection for Word Lists

Philippe Roussille
(Philippe.Roussille@irit.fr)
IRIT, University of Toulouse, France

Abstract

Our work focuses on the selection of words in a list, in the context of visual impairment on mobile devices, without a priori knowledge of the different words. Indeed, most entry systems offer assistance through prediction or correction lists. These lists can easily be used by visually impaired users in a short time. Here, we present three studies to find the interaction which minimizes the time required to select a word from a list without knowing its elements a priori. First, we explored the various possible arrangements, and concluded that a linear spatial arrangement is preferable. Then we studied two kinds of audio feedback, and we determined that audio feedback during the selection is enough. Finally, we proposed four validation actions to retain a validation based on the principle of "*press-release*".

Key words

Selection, visually impaired users, text-input, word, lists.

1. Introduction

Mobile devices are becoming smaller while written communication grows more and more in our daily lives. Faced with the lack of buttons on these devices, using a soft keyboard is one of the possible solutions for easy text-input. It is the lack of tactile cues and the heavy reliance on visual feedback of such devices that make their daily use difficult, if not impossible, for visually impaired users. It is possible for a user to know the keyboards layouts through typing experience, as the layout is static. In turn, this implicit knowledge can be used by a text entry system for visually impaired users[1].

However, dynamic word lists proposed by these keyboards make visually impaired users further penalized when used. Their learning is difficult (if not impossible), and exploration force the user into situations saturated by excessive audio feedback; or by a complete lack of any kind of feedback. Following a preliminary study on the prediction system installed by default on Android, we chose to use lists of four to six items to be representative of daily use situations.

We will try to find an effective solution that allows the user to select an item using the capabilities of a smartphone using only a touch device and an audio feedback.

The problem is threefold: items layout (spatial linear, expanded, etc.), location and type of audio feedback (during, after, voice, etc.), and validation of the user selection (double tap, release, etc.)..

2. Related works

2.1 Item layout

According to Yalla[2] who conducted a comprehensive study of existing solutions, including menus or lists based on audio support techniques (*auditory menus*), there are significant gaps in the recommendations and guidelines for designing interfaces based on audio feedback. From his studies on various provisions, it is preferable to propose a hierarchical layouts to facilitate learning. This approach is, however, impossible for us given the extremely volatile nature of the content: it seems complicated to establish a simple hierarchy that the user could easily hold without causing too much mental stress.

On modern screen readers, there are different strategies to avoid these problems. According to Borodin[3], the majority of current screen readers first show the number of available items and then propose an auxiliary window to navigate through the list (such as JAWS screen reader). The main approach is a complete list of vocalization and a change in the interaction with a "dual-window" approach, which takes time and is not really usable in a mobile environment due to the reduced size of the device.

2.2 Audio feedback

Brewster[4] indicates that it is preferable to decrease the maximum number of terms for interactions on small devices such as smartphones or connected watches. Indeed, according to its results, it is better to increase the number of usable gestures than to add audio or tactile feedback to guide the user to an item to confirm.

On a different approach, Zhao[5] presents the Earpod, a system specifically designed to interact in a list or menu. Every interaction performed by the user produces a sound, while making audio feedback more accurate than visual feedback. This system, according to Zhao et al, becomes less effective when menus are becoming more intertwined or they present many unknown elements to the user.

Finally, when there is a large number of elements, but are easily distinguishable, the use of *earcons* (according to Helle[6]) seems to be the most suitable solution: a earcon, like the icon analogy, is based on a distinctive sound to help find information quickly (for example, an alarm sound for an error, or the sound of crumpled paper to signal suppression).

The study by Walker[7] compares earcons and text-to-speech. According to their results, the speech is more effective when coupled with earcons or spearcons. The spearcons, like the earcons are achieved by increasing the speed of voice synthesis to the point of not being able to be fully recognized. Designers must be careful however: the methods of sonication and earcons are often considered as distraction factors, and therefore are not always appreciated by the majority of users. Moreover, as the prediction lists are not constant, associate a different sound for each word would lead to a much greater cognitive load for users.

2.3 Validation gesture

Validation must, according to Kane et al[8] be solved by an approach using several fingers. In their study, several interactions were tested (for example, search for a contact in a list, browse a music playlist, etc.), focusing on approaches based on the "swipe"; in which the selection is done through one finger while the validation is done by a second finger. This approach simplifies exploration while providing an easy selection to the user. For Kane, it is vital to use a multi-finger approach to facilitate accessibility to the visually impaired.

For Wilson et al[9], an effective solution to select is based on pressure detection. Two validation strategies were studied: leave a finger on an item for a specific time or release the finger. According to their studies, keeping the finger for some time prevents errors, but induced a longer interaction.

As we have seen in our review of the existing works, it is difficult to find a simple solution to quickly select a word from a list whose elements are not known beforehand. As such, we will analyze in detail each part of the selection process: determine the usable layout, know when the voice feedback is to be used, and consider different validation gestures.

3. First study : layout determination

3.1 Interactions

During our first experimental study, we wanted to know what layout would best meet the needs of our users. Therefore, we designed five different layouts. These layouts are grouped into three different categories, sorted by the option we chose to present the interaction for the user: either the user can move his finger on the screen and use the spatial arrangement to move towards an item in particular (for example, he moves his finger over an ordered list from right to left); or the user is based on a relative layout (for example, the further it moves away from a point, the further it goes through the list); or the user has no reliance on any surface and runs through the list indirectly (for example, by swiping from one element to another).

3.1.1 Spatial layout

The first category focuses on spatial approaches. In this category, the elements, all of which are of the same size, are arranged spatially to cover the entire screen. Three interactions are included in this category: interaction along a *line* (the elements are placed on a horizontal line), interaction along a *column* (the elements are arranged in a vertical column), and interaction along a *grid* (the elements are arranged in several columns containing an even number of elements). We called this interaction the spatial interaction.

3.1.2 Dynamic layout

The second type of interaction describes a more dynamic arrangement. It is based on a type selection "press-drag" and validates through release. Unlike a purely spatial approach, the elements are dynamically linked to the position of the user's finger. The first element is positioned where the user first places his finger, and the following items are browsed when the user moves away from that point. The user can confirm the item he is looking for by releasing his finger from the screen. We called this interaction the dynamic interaction.

3.1.3 List layout

This latter category includes an interaction where words are not positioned spatially. Instead, the user can browse the list by sliding his finger left or right, and validates by sliding down. We called this interaction the list interaction.

3.2 Method

The experiment was carried out in two sessions, one for each list size. Each session was divided into five blocks, one for each interaction type (including the spatial variations). The blocks were presented in a different order (counterbalancing guaranteed by a Latin square). Before each block, a learning phase was proposed to the user to allow him to become familiar with the interaction in question. At the end of each session, the subject had to complete a SUS questionnaire for each interaction type, and indicate its preference. The items he had to select through the experiment were randomly selected from fruit names. The subject had to, for each task, find a particular item in a list as quickly and accurately as possible. The item itself could be in any position in the list, or be missing. Every task was repeated three times by word position for each block, for a total of 15 tasks for blocks of four elements, and 21 tasks for blocks of six elements. To avoid learning effect, we randomized the order of the positions in a single block.

3.3 Material and participants

We asked 12 subjects (3 women and 9 men, mean age 28.5 years). All participants were blindfolded to ensure they could not see the phone screen. The device used was a smartphone Samsung Galaxy SII, with a resolution of 306 dpi for a screen of 136.6 mm × 70.6 mm, using a basic Android system, version 4.3. The device was used in landscape mode. The subjects used their fingers to navigate and interact with the elements. The voice and feedback was recorded before the start of the experiment by using Google Translate, then compressed to enable immediate audio feedback and limit possible delays. Nothing was on the screen during the experiment, making the items entierly invisible for users.

3.4 Collected data

In order to study the different interactions, we measured the time taken by the user to make a selection, the number of items browsed, the distance (in pixels) covered, the errors made and the answers to the SUS questionnaires. We also noted any comments the users made. It lead to a total of 2,414 trials.

3.5 Results

Our first criterion for establishing a performance index of the layout in question is the time taken by the user to select an item (from item selection to validation). The second criterion is the number of items browsed by the user before he makes his selection. The third criterion is the distance which corresponds to total length of the path made by the user gestures on the device screen (between a press and a successive release). The fourth criterion we used is the error rate, corresponding to the ratio of elements selected by mistake by the user.

	Column	**Dynamic**	**Grid**	**Line**	**List**
Time (s)	6.1	6.7	6.1	6.2	6.9
Items	3.0	3.1	2.8	2.8	1.8
Distance (px)	424.2	459.3	966.9	683.7	2,737.7
Error	2%	3%	1%	2%	1%

Fig. 1 – Results for 4-items lists

	Column	**Dynamic**	**Grid**	**Line**	**List**
Time (s)	7.5	9.9	7.4	7.4	8.3
Items	4.3	6.0	3.9	4.3	2.8

Distance (px)	467.6	772.3	1,125.1	780.7	3,511.4
Error	4%	10%	3%	4%	3%

Fig. 2 – Results for 6-items lists

3.5.1 Time spent

For the entire article, we will use $p < 0.05$ to indicate a significative result with a p-value smaller than 5 %.

Here, the longest interactions are the list and the dynamic interaction, regardless of the size of the list ($p < 0.05$). The fastest interaction is the grid interaction. We can also notice that the dynamic interaction takes longer as the number of elements increases (6.7 s on average for lists with 4 items versus 9.9 seconds for lists having 6 elements). This is because of a bad starting position : if the user begins near an edge of the screen, the size of items to display is heavily reduced.

3.5.2 Items viewed

The dynamic interaction is the interaction which yields the largest number of items browsed before the user chooses (3.1 items on average for a list of four elements versus 6.0 items for a list of 6 items); while the list interaction is the most efficient one (1.8 items browsed on average items for a list of four items versus 2.8 items for a list of six).

3.5.3 Distance travelled

Here, we can see that the list interaction is the one which produces the greatest distance (2,737 pixels for a list of four elements versus 3,511 for a list of six elements, $p < 0.05$). Conversely, it is the dynamic interaction (500 pixels) which is the most economical when used with four elements ($p < 0.05$); while it is the grid (450 pixels) for six elements ($p < 0.04$).

The disparity between the travelled distance relatively to the list and the low number of items covered is explained by the fact that sliding gestures occupy the entire screen while the actions of the dynamic interaction require much more precision.

3.5.4 Error rate

Grid and list interactions are both offering a minimal error rate (1% for 4 elements and 3% for 6 elements). In contrast, the dynamic interaction produces the most errors for lists of four elements (about 3%), and number even triples for six elements (10%), due to its heavy dependence on precise motion.

3.5.5 SUS questionnaire

Subjects gave the "usable" rating for all interactions when there were only four elements: in fact, every interaction obtained a score of at least 70. The grid interaction was the one preferred. For interactions involving six elements, scores were lower, showing that the dynamic interaction was unusable in this case (it scored 58). This reflects the preferential use: users prefer to use our interactions for smaller lists, and if possible by using a spatial layout (with a subjective preference for the line and the grid approaches).

3.6 Conclusion

Accuracy and time wise speaking, the dynamic interaction is inadequate: it’s too dependent on the starting point (and will fails if the finger is too near the edge of the screen, for example), it requires to proceed by trial and error, using slow motions, browses through more items than other interactions in order to find the requested one. Conversely, the list interaction is much less expensive in terms of accuracy and elements. In return, it needs to repeat a similar browsing gestures (swipe from left or right) which is costly in time and distance.

An interaction based on a spatial layout is thus a good compromise: information can be directly accessed through a finger press, minimizing the distance and the necessary movements. Due to the landscape orientation, the subjects indicated that the column was highly unproductive compared to the line or to the grid. So it seems that a spatial interaction, particularly a linear or grid layout, is best suited.

4. Second study: Importance of auditory feedback and spatial positionning

Based on previous results, we wanted to know whether there was a subjective preference between the grid and the line layouts. Taking into account the time criteria, the number of elements browsed, we checked whether it would possible to determine if one is more effective than the other.

Regarding the auditory feedback, we were offering auditory feedback only while browsing the list. We want to know if it is more effective to offer a complete reading of the list to the user before he starts his choice in order to be faster; or if on the contrary, it is preferable to provide a feedback only during the selection.

We have decided to carry on with the results of the first experiment, comparing only the line and grid layout. As the previous results gave no difference between interactions with four elements and interactions with six elements that were dependent on anything but the length of the list, we focused on a list to four elements.

As there was no significant difference between the distance traveled between the line and the grid, we chose not to use this measure again. Instead, we completed the user feedback through a NASA-TLX questionnaire.

Four prototypes were designed: grid-based selection grid with list reading, grid-based selection grid without preliminary list reading, line-based selection grid with list reading and line-based selection grid without preliminary list reading.

4.1 Method

For our second experience, we asked users to test the four combinations of our parameters (feedback before and during the selection vs. feedback during selection only, using a line vs. using a grid). The order in which we ran the experiment was arranged according to a Latin square to ensure counterbalancement. Before each combination, a learning phase was proposed to the user to allow him to become familiar with the interaction in question.

The requested instructions were the same as those of the first experiment, subjects had to test each interaction technique and find such a component accurately and quickly as possible. Each possible position of element (including lack thereof) has been tested three times. In the same way as the first experiment, the positions were given in random order for each user.

As there was no difference between the interactions we wanted to test, subjects effectuèrent each combination as a unit block of experience. At the end of each block, users had to fill out a SUS questionnaire and a NASA-TLX stress test.

4.2 Participants and apparatus

For this second study, we had 12 subjects (3 women and 9 men, mean age 25 years). All participants were blindfolded to ensure they could not see the phone screen. The device and the experiment conditions were identical to those of the first study.

4.3 Collected data

As we wanted to measure the different interactions, we collected various data during the experiment : the number of items browsed by the user, the time taken by a user to make a selection, the number of errors made, and responses to the two questionnaires. We also collected user feedback.

4.4 Results

We chose to show only the averages of the results obtained under the different conditions (descriptive feedback only, descriptive and prescriptive feedback, line and grid).

	Time (s)	Items browsed	Error rate
Descriptive feedback (DF)	5.3	3.7	2%
Descriptive and prior feedback (PF)	5.6	2.0	2%
Line (L)	5.4	2.8	2%
Grid (G)	5.5	2.9	2%

Figure 3 – Results for both conditions

4.4.1 Time taken by the user

There is no significant impact (p> 0.05) of the type of feedback or layout style on the time taken by the user to choose an item.

4.4.2 Items browsed

As expected, offering to the user the exploration of the list without prior information will make him go through more elements than if he had a prior feedback. However, there is no significant influence of the layout (p > 0.05) on the number of items covered.

4.4.3 Error rate

There is no significant difference (p> 0.05) between the use of prior feedback and its absence, nor of the layout format.

4.4.4 SUS and NASA-TLX questionnaires

We show here that the user answers to the two questionnaires.

	DF	PF	L	G
SUS	76	82	81	76
Mental demand	15%	11%	10%	16%
Physical demand	6%	6%	7%	6%
Temporal demand	8%	11%	10%	9%
Failure rate	13%	13%	10%	13%
Effort	14%	13%	13%	14%
Discouragement	12%	13%	8%	16%

Figure 4 – Qualitative results

4.4.4.1 SUS

Users give a better score when the audio feedback is given prior (76 vs 82, $p > 0.05$), hence stressing the need for voice feedback preceding and accompanying actions. According to users, as long as it is accurate and efficient, slow audio feedback is not detrimental. They preferred a line to a grid layout.

4.4.4.1 NASA-TLX

Users felt more mentally forced when the audio feedback was only descriptive. For them, the grid interaction required more mental effort than the line interaction. Such interaction was more likely to discourage them to answer.

4.5 Conclusion of the second study

For users, a linear interaction seems more relevant than an interaction using a grid: the grid pushes them to use two fingers to move around the screen in a way that they don't see as natural.

In quantitative terms, we see no significant difference for the error rate or the number of browsed elements. In addition, the use of prior feedback, according to some users, can be an obstacle: it is better to browse the list directly for a faster return audio.

Thus it would seem that an ideal interaction presents an linear layout and without prior audio feedback.

5. Third study : Validation gesture

For our third and final study, we proposed four different gestures to validate an item.

5.1 Four selection gestures

We designed four different gestures which can be used to select an item from a list of four elements arranged in a line. We retain, according to our observations made through the two previous studies, one line containing four same size elements, without audio feedback.

In the first three actions proposed here, the user slides his finger on the screen to select an option (selected by slide).

The first validation gesture we propose is based on the principle of "drag and release": the user releases his finger to confirm his choice. We called this gesture "**drag and release**".

The second gesture is based on simultaneous validation using two fingers: the user validates by pressing two fingers simultaneously on the screen. We called this gesture "**dual tap**".

The third gesture performs a validation by a "double tap" (similar to a double-click) the item. We called this gesture "**double tap**".

The fourth and last gesutre is based on the path traced by the finger on the screen. The selection is made by drawing a straight line. As long as the user remains on the line, the system goes through the list. To validate his selection, the user must perform a 90° angle on his route. We called this interaction "**broken line**".

5.2 Participants and apparatus

For this study, we asked 14 subjects with normal vision and a volunteer who was legally blind (6 women and 9 men, mean age 32 years). All participants were all blindfolded to ensure they could not see the phone screen.

The device and the experiment conditions were identical to those of the previous studies.

5.2.1 Procedure

Each participant had to go through the four gestures (the order was counterbalanced by a Latin square). Every gesture made one experiment block, preceded by a training phase. The unitary task, "find the item from the list as quickly and accurately as possible" was repeated twice for each position for a total of eight times per block.

The positions were given in random order to avoid learning effects. At the end of each block, the users had to answer a SUS questionnaire and a NASA-TLX test, as well as rank the different gestures.

5.2.2 Collected data

In order to study the different validation gestures, we have collected various data during the experiment: the number of items browsed by the user, the time taken for the user to make his selection, the errors made and the answers to the SUS questionnaires and NASA-TLX tests. We collected 480 trials.

5.3 Results

Here are the quantitative results we collected.

	Broken line	**Dual tap**	**Drag and release**	**Double tap**
Time (s)	6.8	5.7	5.6	6.2
Items browsed	3.0	2.8	2.7	2.9
Error rate	9%	5%	5%	6%

5.3.1 Time spent to validate an item

There was no significant difference between "dual tap" and "drag-release" (p> 0.05) nor between "broken line" and "double tap" (p> 0.05). However, there is a significant difference between the groups "dual tap"/"drag and release" and "double tap"/"broken line" (p <0.01).

5.3.2 Items browsed before validation

Here as well there is no significant difference between "dual tap" and "drag-release" (p > 0.05) nor between "broken line" and "double tap" (p > 0.05). There is however a significant difference between the two groups "dual tap" / "drag and release" and "double tap" / "broken line" (p < 0.04).

5.3.3 Error rate

There is no significant difference between the error rate for validation "dual tap" / "slide release" (p = 0.34). "Broken line" is less precise than the others (p <0.05).

5.3.4 SUS and NASA-TLX

The results presented here concern the two tests, the SUS questionnaire and NASA-TLX questionnaire.

	Drag and release	**Double tap**	**Broken line**	**Dual tap**
SUS	85	83	68	72
Mental demand	1%	1%	19%	5%
Physical demand	2%	3%	2%	6%
Temporal demand	3%	1%	4%	4%
Failure rate	3%	18%	24%	8%
Effort	3%	4%	8%	4%
Discouragement	3%	9%	22%	14%

Figure 6 – Qualitative results

5.3.4.1 SUS scores

Based on user responses, the "broken line" solution is not satisfactory (SUS scores below 70) while "double tap" and "drag and release" have been validated as favorites interactions (with scores of 83 and 85).

5.3.4.2 NASA-TLX

Users are not felt tired or discouraged by gestures, if not by the "broken line". According to them, it is the interaction inducing the most errors and which is the most discouraging.

5.3.4.3 Ranking

The user ranking indicates a strong preference for the "drag and release" followed by "double tap", "broken line" and finally "dual tap".

5.4 Conclusion of the third study

Timed validation (by a double tap) or drawn validation (drawing lines) are longer and take more time to get the same accuracy as methods using direct validation as a dual tap. These slower methods are thus more precise for the user, but discouraging. Therefore, the interaction drag-release is the best to meet our criteria.

6 General conclusion

In this article, we stress the importance of a simple approach to design lists suitable for selection and validation for visually impaired users in contexts allowing no learning or memorization. Information must be directly accessible, preferably arranged linearly (or on a grid if there is too many elements). The audio feedback is only required when selecting elements (no significant gain in terms of time). Finally, validation must be fast and accurate, if possible in one simple gesture.

Future work

We did restrict ourselves to a list of words. But, in the case of text input, a possible improvement would be to analyze the context, and sort the list of suggestions or corrections to faster and more accurate. We see this technique as a result of our ongoing work to improve the efficiency of our system.

Finally, the work we present is a proof of concept with only blind users, we plan to integrate it on a longer study with users in a visually impaired situation.

References

1. P. Roussille, M. Raynal, and C. Jouffrais, “DUCK : a deDUCtive soft Keyboard for visually impaired users,” in 27ème conférence francophone sur l’Interaction Homme-Machine. Toulouse, France : ACM, Oct. 2015, p. a19.

2. P. Yalla and B. N. Walker, “Advanced Auditory Menus : Design And Evaluation of Auditory Scroll Bars,” Assets’08, pp. 105–112, 2008.
3. Y. Borodin, J. P. Bigham, G. Dausch, and I. V. Ramakrishnan, “More than meets the eye : A survey of screen-reader browsing strategies,” in Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A) , ser. W4A ’10.
4. S. Brewster, J. Lumsden, M. Bell, M. Hall, and S. Tasker, “Multimodal ’eyes-free’ interaction techniques for wearable devices,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI ’03.
5. S. Zhao, P. Dragicevic, M. Chignell, R. Balakrishnan, and P. Baudisch, “Earpod : Eyes-free menu selection using touch input and reactive audio feedback,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems , ser. CHI ’07.
6. S. Helle, G. LePlâtre, J. Marila, and P. Laine, “Menu sonification in a mobile phone – A prototype study.” pp. 255–260, 2001.
7. B. N. Walker and A. Kogan, “Spearcon performance and preference for auditory menus on a mobile phone,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) , vol. 5615 LNCS, no. PART 2, pp. 445–454, 2009.
8. S. K. Kane, J. P. Bigham, and J. O. Wobbrock, “Slide rule : Making mobile touch screens accessible to blind people using multi-touch interaction techniques,” in Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility, ser. Assets ’08.
9. G. Wilson and S. Brewster, “Pressure-Based Menu Selection for Mobile Devices,” 2010, pp. 181–190.