SPINDEX AND SPEARCONS IN MANDARIN: AUDITORY MENU ENHANCEMENTS SUCCESSFUL IN A TONAL LANGUAGE

Thomas M. Gable

Sonification Lab,
Georgia Institute of Technology,
654 Cherry Street, Atlanta, GA, 30332, USA
thomas.gable@gatech.edu

Brianna Tomlinson

Sonification Lab,
Georgia Institute of Technology,
654 Cherry Street, Atlanta, GA, 30332, USA
btomlin@gatech.edu

Stanley Cantrell

Sonification Lab,
Georgia Institute of Technology,
654 Cherry Street, Atlanta, GA, 30332, USA
cantrell@gatech.edu

Bruce N. Walker

Sonification Lab,
Georgia Institute of Technology,
654 Cherry Street, Atlanta, GA, 30332, USA
bruce.walker@psych.gatech.edu

1. ABSTRACT

Auditory displays have been used extensively to enhance visual menus across diverse settings for various reasons. While standard auditory displays can be effective and help users across these settings, standard auditory displays often consist of text to speech cues, which can be time intensive to use. Advanced auditory cues including spindex and spearcon cues have been developed to help address this slow feedback issue. While these cues are most often used in English, they have also been applied to other languages, but research on using them in tonal languages, which may affect the ability to use them, is lacking. The current research investigated the use of spindex and spearcon cues in Mandarin, to determine their effectiveness in a tonal language. The results suggest that the cues can be effectively applied and used in a tonal language by untrained novices. This opens the door to future use of the cues in languages that reach a large portion of the world's population.

2. INTRODUCTION

The modern computer interface primarily relies on the use of the visual modality to relay information to the user. However, the employment of visuals is not always viable for users, whether it is due to vision impairment, lack of visual clarity from a system because of increasingly smaller screens, or situational blindness such as when the user is driving or completing another visually-demanding task. In these instances the transfer of information to the user via auditory displays can often be a viable option. Examples of auditory displays range from fire alarms and other basic interfaces to complex computer systems. These systems can employ speech-based or non-speech audio to relay information to users. Extensive research has shown the successful deployment of auditory displays for user interfaces ranging from warning systems to menus.

In the current document we explore the ability of novices to use two popular advanced auditory cues (spindex and spearcons) in Mandarin, a tonal language in which these advanced cues have not yet been extensively evaluated.

3. AUDITORY MENUS

Menus are a part of many interfaces that we interact with on a daily basis. Their complex structure and the difficulty in navigating through them while understanding what they contain can pose a large problem in the employment of auditory interfaces. Individuals with vision impairment rely on auditory menus and displays to interact with user interfaces for computers, phones, and other technologies; these devices rely on screen readers, which use text-to-speech (TTS) output to provide contextual description for the software content. The TTS interfaces employed in these screen readers are one of the simplest forms of auditory feedback from a computer interface and have been used for many types of auditory displays to support accessibility. In addition, auditory menus have become widely used in hands-free calling or other contexts when a user may be situationally visually impaired, such as when driving or doing other complex visual-manual tasks [1].

3.1. Speech-based interfaces

These speech-based interfaces have been shown to relay information to the users fairly well. However, TTS and other interfaces are limited in their abilities to provide a fast interaction and cannot always relay all details to the users successfully. Previous work has shown that TTS interfaces can lead to slower interaction times than visual interaction [2] [3]. One study found that in a dual-task situation when users were asked to complete a search task while driving they employed the TTS auditory menus until the importance of the secondary task was heightened, at which point the users abandoned the auditory display and relied on visuals to

https://doi.org/10.21785/icad2017.025

complete the task [4]. These findings show that when users are given a secondary task rated with a higher importance than other tasks (and have the ability to choose interface modality), they will choose to use their visual attention to complete the task if the interaction time is too long using an auditory displays, even if it leads to more dangerous primary task behaviors.

3.2. Non-speech interfaces

In order to address the issue of slow feedback in TTS interfaces, researchers have developed a number of non-speech auditory approaches to transfer information to the user, including auditory icons, earcons, and audemes. Auditory icons are sounds which portray real world concepts and objects directly and usually have a one to one mapping, such as the sound of a *creaking door* representing *entering a new room* or *opening a file* [5]. Auditory icons leverage previous knowledge on the part of the listener, and provide an easy way to transfer knowledge to the designed sound. Earcons are another type of non-speech auditory display, but instead of mimicking or directly representing real-world sounds, they are purposefully designed, musical tones [6]. Audemes present a more complex representation through layers of multiple sets of music or sound effects to portray more complex themes or concepts to support educational contexts [7]. Though these types of displays can be easily recognizable and learnable, they can potentially lead to large amount of memorization for users, and are not as usable for supporting menu or long lists of options.

3.3. Advanced auditory cues

To support faster navigation of auditory menus compared to typical TTS menus, several types of advanced auditory cues have been created, including spearcons and spindex. *Spearcons* are unique, algorithmically-condensed pieces of speech, which are based on their original TTS words or phrases, but are often no longer recognizable as a specific spoken utterance [8] [9]. Previous work has found that spearcons reduce learning time for an auditory menu compared to other non-text auditory feedback such as earcons [10]. *Spindex* cues are a set of sounds that represent the initial sounds of menu items, to support faster searching through a large auditory list or menu (e.g., a listener might use the spindex to skip past As, Bs, Cs, and so on, until they reach T, which they know is the first letter of a song title) [11]. Spindex cues have been shown to decrease search time compared to plain TTS; and users report them to have a high level of helpfulness compared to other speech-based auditory menus [11][12].

Though spindex and spearcons were both initially developed in English, some researchers have explored using spearcons and spindex to enhance TTS interfaces across multiple languages including German, Hungarian, Korean, and Swedish [13][12][14]. Recently, researchers have evaluated how well Mandarin spearcons support eyes-free navigation of real physical environments (i.e., not used in a user interface) compared to English spearcons and to plain Mandarin TTS (with no enhancements) [15]. They found that Mandarin spearcons are better at conveying distance from a target, but the methodology of that study relied heavily on training for both types of spearcons to scaffold initial learning, and had a limited set of spearcons that the listeners used to navigate. Something which was not shown was whether or not spearcons are successful for user interfaces in tonal languages (such as Mandarin) without this extensive training and with a larger set of cues. Spearcon-shortening of speech sounds might affect the tonal pronunciation for a broad range of words, like those which might show up in a long playlist for songs. In addition, there has not been extensive exploration of spindex cues in languages other than English, particularly in tonal languages. We explore both of these concepts through our two experiments below.

4. CURRENT STUDY

The current studies investigated the ability of spindex and spearcon cues to work in a tonal language for long list and multiple tab menu navigation with no training. Two separate studies were undertaken: one for menu navigation of long lists, and another for navigation of multiple-tab menus.

It was expected that both auditory cues would work in Mandarin, be recognizable by participants, and assist in navigation without any training necessary as seen through either a lack of any differences in performance between the TTS and advanced auditory cues, or seeing better performance with the advanced auditory cues.

5. STUDY 1

5.1. Participants

A total of 23 participants (15 males and 8 females) with an average age of 22.1 (SD = 4.5) from a large research university in the United States took part in the study. Only native speakers of standard Mandarin were recruited for the study, and all participants were required to have normal or corrected to normal vision and hearing. Participants reported having spoken Mandarin for an average of 20.4 years (SD=5.9) and writing the language for an average of 18.2 years (SD=5.5). The mean self-reported fluency ratings for Mandarin on a scale from 1-6 (6 being the highest fluency) was 5.9 (SD=0.3), and for writing it was 5.7 (SD=0.9). All participants signed informed consent and provided demographic information; they also completed a few questions regarding their knowledge of Mandarin to ensure an equivalent minimum level of Mandarin expertise.

5.2. Apparatus

Visual stimuli were presented using a 21.5 inch monitor with 1440 x 900 pixel resolution; auditory stimuli were presented using Sony MDR-7506 Studio headphones. Participant responses were collected in a sound attenuated chamber to isolate the sounds used in the research, and to ensure no other noises competed for the participants' attention. A software program written in JAVA and using the APWidgets library [16] was run from an iMac computer, with a 2GHz Intel Core 2 Duo processor and 1GB of RAM running Mac OSX 10.10.4 and displayed on the 21.5 inch monitor described above. The software was created to randomize across a block system, cue participants to when a task needed to be completed, collect responses, and record data. Participants used the connected keyboard to input their responses to the computer, which was placed on a desk in front of them.

To measure subjective workload the common measure of NASA Task Load Index (NASA-TLX) was used. The index measures six subscales of effort including effort, temporal

demand, physical demand, frustration, performance, and mental demand [17].

5.3. Menu structures

The song list menu in this study was created by taking SBS POP ASIA 2015 top 100 songs in China and removing titles that were in English, or translating the English song titles to Mandarin, as appropriate. This left 94 songs in total, which were then sorted alphabetically using Mandarin pinyin ordering. The menu was navigated using the arrow keys on the keyboard, selecting a song with the enter key.

5.4. Auditory stimuli

The auditory stimuli were created by recording a female native Mandarin speaker reading out all of the menu items used in the study. This was done due to the text-to-speech (TTS) generators currently available on the market being rated poorly and being said to sound "unnatural" by native speakers in an initial pilot. These human recordings were then put through a number of algorithms to create the necessary spearcon and spindex cues for the menu system. The spearcon cues were created by a C++ program used to create a linear logarithmic compression of the TTS audio (.WAV) files. Spindex cues for the study were determined by the pinyin of each song item and each unique pinyin was recorded as individual audio (.WAV) files from the female native Mandarin speaker. These audio files were then stitched together using Sound eXchange to create the more complex auditory cues [18].

5.5. Procedure

Upon arrival participants were asked to confirm they met the study criteria and then read through and signed an informed consent form. Then, the participants were assigned to the order of which menu structure they completed first, which was randomized between participants.

Participants were then introduced to the structure of the menu by looking at and interacting with it visually. No audio was provided as they did this, and participants used the arrow keys to move up and down the list, as practice for the study block. Participants were also able to see where the target item would be shown during the experiment. Next, the participants started the condition blocks; these consisted of a training phase of 5 random item selections followed by 20 items for each condition. These 20 items were selected via a semi-random bin system where the total list was divided into 4 bins and one song was pulled from each bin before repeating a bin. This was done to ensure choices throughout the entire list instead of the possibility of an uneven distribution. There were 5 conditions in the study including Visual-only, TTS, Spindex+TTS, Spearcon+TTS, and Spindex+Spearcon+TTS. After each condition participants completed the NASA-TLX to measure subjective workload and moved onto the next condition in the block. In total there were 3 blocks of each of the 5 conditions, resulting in 15 total blocks.

After completing all 15 blocks of menu search tasks, participants completed a demographics survey about preferences for the cues, perceived fun of use, likability, annoyance, helpfulness, and effectiveness of the cues.

5.6. Analysis

The data for time to complete each selection task, accuracy, and TLX scores were collected for each block. All of these data were analyzed with within-subject repeated measures ANOVA, with Huynh-Feldt corrections, as appropriate, for violations of sphericity assumptions. The survey data were analyzed via basic paired t-tests.

5.7. Results

5.7.1. Interaction (Search) Time

The data for interaction (search) time are shown in Figure 1 (broken into conditions to allow for more detail of the conditions). A two way repeated measures ANOVA with Huynh-Feldt corrections was done on the interaction time results, and revealed a significant main effect of block, $F(1.67, 35.05) = 109.98$, $p < .001$. To determine the differences in interaction time between blocks, post-hoc comparisons of 3 t-tests with Bonferroni corrections (correcting alpha to .0166) were performed. The post-hocs revealed significant differences between Blocks 1 and 2, $t(22) = 10.41$, $p < .001$; Blocks 1 and 3, $t(21) = 12.42$, $p < .001$; and Blocks 2 and 3, $t(21) = 4.39$, $p < .001$.

The original two-way repeated measures ANOVA found no significant main effect of condition $F(3.98, 83.51) = 0.96$, $p = .435$, and no significant interaction, $F(5.55, 116.60) = 1.49$, $p = .193$.

These results show a practice effect, in that participants were significantly faster in each subsequent block, across the three conditions.

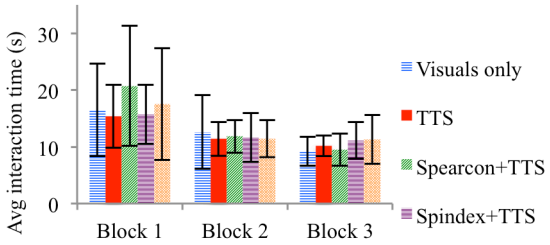

Figure 1: Average interaction time across the three blocks for the 5 conditions. Note that there was a significant difference between the three blocks across the conditions.

5.7.2. Accuracy

A two way repeated measures ANOVA (Huynh-Feldt corrections) was performed on the accuracy results and found no significant main effect of condition, $F(2.95, 61.85) = 0.67$, $p = .570$; no significant main effect of block, $F(1.40, 29.29) = 1.04$, $p = .342$; and no significant interaction, $F(5.51, 115.77) = 0.26$, $p = .944$. This means that accuracy was consistent in the song list experiment across conditions and blocks.

5.7.3. NASA-TLX

Perceived workload (NASA TLX) data are shown in Figure 2. A one way repeated measures ANOVA (Huynh-Feldt corrections) on the NASA-TLX results showed a significant

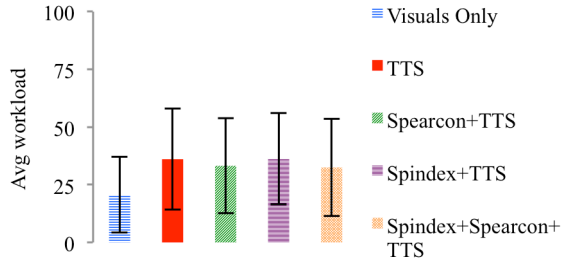

Figure 2: The average subjective workload as rated via NASA-TLX for each condition, averaged across blocks.

main effect of condition $F(2.87, 48.90) = 4.55$, $p = .008$. To determine where the differences between the conditions were taking place, a set of post-hoc analyses were performed via 10 paired t-tests with Bonferroni corrections (correcting alpha to .005). These analyses can be seen in Table 1. The tests revealed significantly less total subjective workload in the Visuals condition than the Spearcon+TTS condition, the Spindex+TTS condition, and the Spindex+Spearcon+TTS condition. No other differences were found to be significant. These results mean that participants reported higher total subjective workload in the Spearcon+TTS, Spindex+TTS, and Spindex+Spearcon+TTS conditions than they did in the Visuals condition.

5.7.4. Survey

The paired t-tests for the survey questions revealed no significant differences between the conditions in regards to preferences.

5.8. Discussion

The results of Study 1 showed participants were no faster or slower at finding the songs for any one condition, but were faster for each block on average across all conditions. These results suggest that the spindex and spearcon cues were able to convey information to the participants effectively with no training and that the extended number of cues was not a problem for participants. These results support the hypothesis that the spindex and spearcon cues could be used to navigate a long list in Mandarin, with no extended practice.

Block x Condition pairs	*t*	*df*	*p*
Visuals – TTS	2.97	17	.009
Visuals – Spearcon+TTS	3.51	17	.003*
Visuals – Spindex+TTS	4.14	17	.001*
Visuals – Spindex+Spearcon+TTS	3.77	17	.002*
TTS – Spearcon+TTS	0.24	17	.816
TTS – Spindex+TTS	0.53	17	.600
TTS – Spindex+Spearcon+TTS	0.42	17	.680
Spearcon+TTS – Spindex+TTS	0.64	17	.530
Spearcon+TTS – Spindex+Spearcon+TTS	0.55	17	.591
Spindex+TTS – Spindex+Spearcon+TTS	0.95	17	.354

Table 1: The paired t-test post hoc results done for TLX data in study 1. Note that * marks a significant difference.

The results also showed that participants had no significant difference in accuracy in finding the song across either condition or blocks. Again, this supports the hypothesis of the cues being able to be used effectively in Mandarin.

Although no time or accuracy differences were found, participants did report higher total workload in the Spearcon+TTS, Spindex+TTS, and Spindex+Spearcon+TTS conditions than they did in the Visuals condition. This hints that the subjective mental demand it took to complete the task was higher for the non-visual conditions, but this effect may decrease with practice.

文件	编辑	查看	收藏夹	工具	帮助
新建选项卡	剪切	工具栏	添加到收藏夹	删除浏览历史记录	Internet Explorer 帮助
重复打开选项卡	复制	浏览器栏	添加到收藏夹栏	InPrivate浏览	Internet Explorer 11中的新功能
新建窗口	黏贴	转到	将当前所有的网页添加到收藏夹	启用跟踪保护	联机支持
新建会话	全选	停止	整理收藏夹	ActiveX筛选	关于Internet Explorer
在沉浸式浏览器中打开	在此页上查找	刷新	Links	修复连接问题	
打开		缩放	Bing	重新打开上次浏览会话	
使用Notepad编辑		文字大小		将站点添加到“应用”视图	
保存		编码		查看下载	
另存为		样式		弹出窗口阻止程序	
关闭选项卡		插入光标预览		SmartScreen筛选器	
关闭所有标签		源		管理媒体许可证	
页面设置		安全报告		管理加载项	
打印		国际网站地址		兼容性视图设置	
打印预览		网页隐私策略		订阅此源	
发送		全屏		源发现	
导入和导出				Windows更新	
属性				性能仪表板	
退出				F12开发人员工具	
				OneNote Linked Notes	
				Send to OneNote	
				Internet选项	

Table 2: The Internet Explorer menu layout used in Study 2.

6. STUDY 2

6.1. Participants

A total of 23 participants (12 males and 11 females) with an average age of 20.6 (SD = 2.3) from the same university in the United States took part in the study. Again, only native speakers of Mandarin were recruited for the study and all participants were required to have normal or corrected to normal vision and hearing. Participants reported speaking Mandarin for an average of 20.3 years (SD=2.4) and writing the language for an average of 16.2 years (SD=4.9). The mean self-reported fluency ratings for Mandarin on a scale from 1-6 (6 being the highest fluency) was 5.9 (*SD*=0.3), and for writing it was 5.7 (*SD*=1.1). All participants signed informed consent and provided demographic information, and completed a few questions regarding their knowledge of Mandarin to ensure an equivalent minimum level of Mandarin expertise.

6.2. Apparatus

The apparatus for Study 2 was the same used in Study 1.

6.3. Menu structures

The menu system for Study 2 was based on the set of menu options available on Internet Explorer version 11, with one additional option ("Close all tabs") to create an even set of blocks to randomize within. This caused the structure of the menu out to have 69 menu items, under 6 tabs. The structure and items for the menu can be seen in Table 2.

6.4. Procedure

As in Study 1, participants were asked to confirm they met the study criteria and then read through and signed an informed consent form upon arrival. Participants were first introduced to the structure of the menu by looking and interacting with it visually. No audio was provided as they did this. During this time participants used the arrow keys to move left, right, up, and down, as they would during the study. They were also shown where the target item would be displayed on the screen once the study began. Following this orientation, participants started the condition blocks; these consisted of a training phase of 5 random item selections followed by 23 item selections (one for each available menu item) for each condition. The three conditions in the main menu blocks were Visual-only, TTS, and Spearcon+TTS; the order of conditions was randomized via a Latin square. Each condition was completed a total of 3 times, resulting in 9 total blocks. After each condition, participants completed the NASA-TLX to measure subjective workload.

Following the completion of all blocks of menu search tasks, participants completed a demographics survey and the same preferences questions as were given in Study 1.

6.5. Analysis

As in Study 1, data for interaction (search) time, accuracy, and TLX scores were all collected for each block. All of these data were analyzed via within subject repeated measures ANOVA with Huynh-Feldt corrections for sphericity. The survey data were analyzed via paired t-tests.

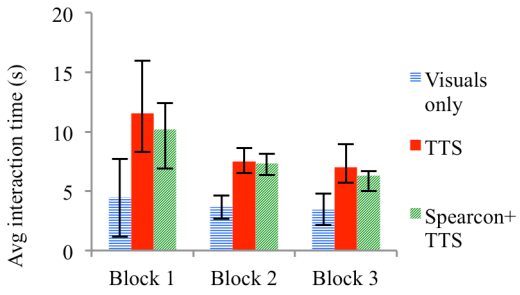

Figure 3: Average interaction time (in seconds) across the 3 blocks and 3 conditions for the IE menu task.

6.6. Results

6.6.1. Interaction time

The data for interaction (search) time across blocks and conditions are shown in Figure 3. A two way repeated measures ANOVA (Huynh-Feldt corrections) was done for interaction time, and revealed a significant main effect of condition $F(1.97, 37.47) = 90.80$, $p < .001$; a significant main effect of block, $F(1.30, 24.75) = 61.26$, $p < .001$; and a significant interaction, $F(1.93, 36.70) = 4.84$, $p = .014$.

To determine the differences in interaction time between conditions, post-hoc comparisons of 3 t-tests with Bonferroni corrections (alpha = .0166) were performed. The post-hocs revealed no significant difference between Spearcon+TTS and TTS, $t(21) = 1.21$, $p = .240$; but did show a significant difference between Spearcon+TTS and Visuals, $t(21) = 11.06$, $p < .001$; and for TTS and Visuals, $t(21) = 12.56$, $p < .001$. These results show that participants were significantly faster during the Visuals condition than both the Spearcon+TTS and the TTS conditions.

To determine the differences in interaction time between blocks, post-hoc comparisons of 3 t-tests with Bonferroni corrections (alpha = .0166) were performed. The post-hocs revealed significant differences between Blocks 1 and 2, $t(21) = 7.69$, $p < .001$; Blocks 1 and 3, $t(20) = 9.25$, $p < .001$; and Blocks 2 and 3, $t(20) = 4.20$, $p < .001$. These results reflect a practice effect, in that participants were significantly faster in each subsequent block.

To determine what interactions between condition and block were happening in the data, two sets of post-hoc comparisons of 9 t-tests (Bonferroni corrections; alpha = .0056) were performed. The first set of post-hocs was done to look at the interaction time differences between conditions within each block. These analyses can be seen in Table 3. The analyses revealed that the Visuals condition was significantly faster than either the Spearcon+TTS condition or the TTS condition within each block.

The second set of post-hocs was done to investigate the interaction time differences between blocks in each condition. The analyses showed that there were significant differences within the Spearcon-TTS and TTS conditions for each block. This means that participants got faster at using the Spearcon-TTS and TTS auditory cues as they progressed through the three blocks, which could be an argument for them learning how to use the cues more efficiently. No such learning effect was seen for the Visual-only condition, suggesting that it was not the learning of the menu that sped up the participant's interactions. The analysis data can be seen in Table 4.

Condition x Block pairs	t	df	p
Block 1 Spearcon+TTS – Block 1 TTS	1.05	21	.307
Block 1 Spearcon+TTS – Block 1 Visuals	6.61	21	< .001*
Block 1 TTS – Block 1 Visuals	6.36	21	< .001*
Block 2 Spearcon+TTS – Block 2 TTS	0.69	21	.495
Block 2 Spearcon+TTS – Block 2 Visuals	14.59	19	< .001*
Block 2 TTS – Block 2 Visuals	13.48	19	< .001*
Block 3 Spearcon+TTS – Block 3 TTS	1.74	20	.098
Block 3 Spearcon+TTS – Block 3 Visuals	9.61	20	< .001*
Block 3 TTS – Block 3 Visuals	8.34	20	< .001*

Table 3: Table of the post hoc analyses performed to investigate interaction time differences between conditions within each block for the IE menu task.

Block x Condition pairs	t	df	p
Block 1 Spearcon+TTS – Block 2 Spearcon+TTS	4.02	21	.001*
Block 1 Spearcon+TTS – Block 3 Spearcon+TTS	6.37	20	< .001*
Block 2 Spearcon+TTS – Block 3 Spearcon+TTS	3.89	20	.001*
Block 1 TTS – Block 2 TTS	4.15	21	< .001*
Block 1 TTS – Block 3 TTS	9.25	20	< .001*
Block 2 TTS – Block 3 TTS	15.32	20	< .001*
Block 1 Visuals – Block 2 Visuals	1.59	19	.128
Block 1 Visuals – Block 3 Visuals	1.96	20	.064
Block 2 Visuals – Block 3 Visuals	1.34	19	.197

Table 4: Table of the post hoc analyses performed to investigate interaction time differences between blocks in each condition

6.6.1. Accuracy

The data are shown in Figure 4. A two-way repeated measures ANOVA (Huynh-Feldt corrections) was done on the accuracy data, and found a main effect of block $F(1.18, 22.38) = 5.00$, $p = .031$. To determine the differences in accuracy between blocks post-hoc comparisons of 3 t-tests with Bonferroni corrections (alpha = .0166) were performed. The post-hocs revealed no significant difference between Blocks 1 and 2, $t(21) = 1.77$, $p = .091$; nor Blocks 1 and 3, $t(20) = 2.42$, $p = .025$; but did show a significant difference between Blocks 2 and 3 for the three conditions, $t(20) = 3.31$, $p = .003$.

The main ANOVA found no significant main effects of condition $F(1.36, 25.89) = .542$, $p = .522$, nor interactions $F(2.17, 41.18) = 1.52$, $p = .230$.

These results show a practice effect nearer the end of the study, in that participants were significantly more accurate across all three conditions in Block 3 than in Block 2.

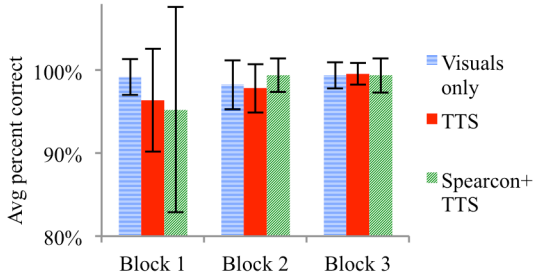

Figure 4: Average accuracy across conditions and blocks in the IE menu task.

6.6.2. NASA-TLX

A one way repeated measures ANOVA (Huynh-Feldt corrections) was done on the NASA-TLX results, and showed no significant main effects for condition $F(1.78, 35.66) = 1.33$, $p = .276$. This means that between the three conditions in the Internet Explorer study participants did not rate the conditions as being any different from each other in perceived workload.

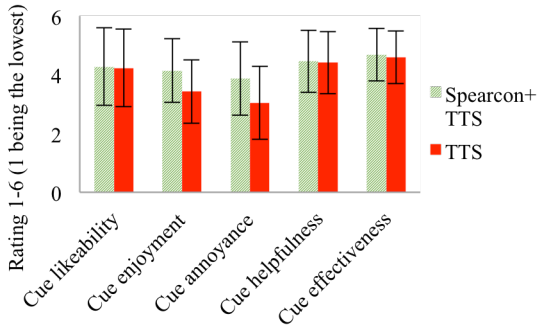

Figure 5: Average preference ratings across the spearcon+TTS and TTS cues in the IE menu study.

6.6.3. Survey

The survey data are shown in Figure 5. Paired samples t-tests on the survey data revealed that participants rated the Spearcon+TTS as more fun than TTS, $t(20) = 2.20$, $p = .040$. The analyses also revealed that participants considered Spearcon+TTS as more annoying than TTS, $t(21) = 2.41$, $p = .025$.

6.7. Discussion

In Study 2 it was found that participants were faster when using the visuals condition than either TTS or spearcon condition. This is expected, and suggests that the auditory-only cues, on the whole, slowed participants down; but the lack of any difference between spearcons and TTS for both speed and accuracy mean that the spearcons were effective at relaying information to the users. When investigating the learning effect on interaction (search) time, the results

showed that participants were getting faster in the two auditory conditions, but not the visuals condition. This suggests that the auditory conditions may have eventually been performed at the same speed as the visual conditions had participants had enough practice.

The results also showed, similarly to Study 1, that there was no significant difference in accuracy across the conditions, suggesting participants had similar ability in finding the correct menu item whether using spearcons, TTS, or visuals. The result of their learning was shown across all three conditions (Block 3 being more accurate than Block 2), but no other differences between conditions suggested similar performance across each condition.

Finally, NASA TLX results showed that participants had no perceived workload difference between the three conditions, meaning that the auditory cues were not subjectively harder to use than the visuals.

7. Overall Discussion

In sum, the results from the two studies suggest that spearcons and spindex cues are effective for use in Mandarin. The lack of any significant differences in accuracy across the Spindex or Spearcon cues and the TTS or visuals conditions suggests that participants were able to use the cues effectively to choose the required items. While time differences were seen between the advanced auditory cues and other cues in the IE task, participants would be expected to get even faster with these cues as they have more practice.

These results are similar to those seen in studies using spearcons in English [10] and Korean [12]. In these studies participants were also able to quickly learn to use the cues to effectively complete the tasks. It was seen in both of these previous studies that after a time of practice participants were actually faster with the spearcon cues. While this was not found in the current study, participants had less trials with the cues in this study so people may be found to be faster with more practice with the Mandarin version of the cues. Another similarity with the current study to previous work was that as with the study don in Korean [12], participants found spearcons to be more fun to use than TTS alone, which suggests people may be willing to use them in the real world.

7.1. Application of Results

These results suggest that the use of spindex and spearcon auditory cues in Mandarin could be effective in visually demanding multi-tasking situations or times when visual displays are not available. Screen readers for blind individuals could use these types of cues in Mandarin and most likely other tonal languages. As has been shown to work in English [2][19] these cues may help drivers to more safely perform secondary tasks in the car such as finding a radio station, or completing other tasks while keeping their eyes on the road.

7.2. Limitations

There were a few factors in the research that some might consider limitations including the use of college students and those who spoke English in addition to Mandarin. However, the use of students or their knowledge of English should not change the participants' abilities to perform the task as a Mandarin speaker. What should be considered is the performance with the auditory cues and the workload associated with them, as compared to the Visuals-only condition. The higher workload reported by the participants could be considered in applications of the work; however, having more practice with the cues would be expected to decrease this workload difference.

8. Conclusions

The results of this research suggest that Mandarin spearcons can work across an extended vocabulary and in multiple settings with no extensive training needed. In addition, the research suggests that spindex in Mandarin can help users move through a list effectively. This implies that these types of auditory cues can be used extensively in even more languages than previously known, and provide a suggestion that the cues can work in other tonal languages as well.

9. ACKNOWLEDGMENT

Portions of the work were supported by National Science Foundation Graduate Research Fellowships (DGE-1148903, DGE-1650044) as well as additional grant funding from the NSF and from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR). We would like to thank Grace Li and Alexis Wilkinson for collecting data for these studies.

10. REFERENCES

[1] Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., & Baudisch, P. (2007, April). Earpod: eyes-free menu selection using touch input and reactive audio feedback. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1395-1404). ACM.

[2] Gable, T. M., Walker, B. N., Moses, H. R., & Chitloor, R. D. (2013, October). Advanced auditory cues on mobile phones help keep drivers' eyes on the road. In Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 66-73). ACM.

[3] Yin, M., & Zhai, S. (2006, April). The benefits of augmenting telephone voice menu navigation with visual browsing and search. In Proceedings of the SIGCHI conference on Human Factors in computing systems (pp. 319-328). ACM.

[4] Brumby, D. P., Davies, S. C., Janssen, C. P., & Grace, J. J. (2011, May). Fast or safe?: how performance objectives determine modality output choices while interacting on the move. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 473-482). ACM.

[5] Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. Human-computer interaction, 2(2), 167-177.

[6] Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. Human–Computer Interaction, 4(1), 11-44.

[7] Ferati, M., Pfaff, M. S., Mannheimer, S., & Bolchini, D. (2012). Audemes at work: Investigating features of non-speech sounds to maximize content recognition.

International Journal of Human-Computer Studies, 70(12), 936-966.

[8] Walker, B. N., Lindsay, J., Nance, A., Nakano, Y., Palladino, D. K., Dingler, T., & Jeon, M. (2013). Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*(1), 157-182.

[9] Walker, B. N., Nance, A., & Lindsay, J. (2006). Spearcons: Speech-based earcons improve navigation performance in auditory menus.

[10] Palladino, D. K., & Walker, B. N. (2007). Learning rates for auditory menus enhanced with spearcons versus earcons.

[11] Jeon, M., & Walker, B. N. (2009, October). "Spindex": Accelerated Initial Speech Sounds Improve Navigation Performance in Auditory Menus. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 53, No. 17, pp. 1081-1085). SAGE Publications.

[12] Suh, H., Jeon, M., & Walker, B. N. (2012, September). Spearcons improve navigation performance and perceived speediness in Korean auditory menus. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 1361-1365). SAGE Publications.

[13] Larsson, P., & Niemand, M. (2015). Using sound to reduce visual distraction from in-vehicle human–machine interfaces. Traffic injury prevention, 16(sup1), S25-S30.

[14] Wersényi, G. (2010). Auditory representations of a graphical user interface for a better human-computer interaction. In Auditory Display (pp. 80-102). Springer Berlin Heidelberg.

[15] Jeon, M., Gable, T. M., Davison, B. K., Nees, M. A., Wilson, J., & Walker, B. N. (2015). Menu navigation with in-vehicle technologies: auditory menu cues improve dual task performance, preference, and workload. *International Journal of Human-Computer Interaction*, *31*(1), 1-16.

[16] Davison, B. K., & Walker, B. N. (2008). AudioPlusWidgets: Bringing sound to software widgets and interface components. Proceedings of ICAD2008, Paris, France.

[17] Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload, 1*(3), 139-183.

[18] SoX - Sound eXchange [Computer Software]. (2015). Retrieved from sox.sourceforge.net.

[19] Hussain, I., Chen, L., Mirza, H. T., Wang, L., Chen, G., & Memon, I. (2016). Chinese-Based Spearcons: Improving Pedestrian Navigation Performance in Eyes-Free Environment. International Journal of Human-Computer Interaction, 32(6), 460-469.