# TOWARDS SCREEN READERS WITH CONCURRENT SPEECH: WHERE TO GO NEXT?

João Guerreiro
INESC-ID, Instituto Superior Técnico, Universidade de Lisboa
joao.p.guerreiro@tecnico.ulisboa.pt

## Introduction

Blind people rely mostly on the auditory feedback of screen readers to consume digital information. Despite the browsing strategies employed by blind users [3], how fast can information be processed remains a major problem. Sighted people use scanning as a strategy to achieve this goal, by glancing at all content expecting to identify information of interest to be subsequently analyzed with further care. In contrast, screen readers rely on a sequential auditory channel that is impairing a quicker overview of the content, when compared to the visual presentation on screen.

We proposed taking advantage of the *Cocktail Party Effect* [6], which states that people are able to focus their attention on a single voice among several conversations, but still identify relevant content in the background. Therefore, oppositely to one sequential speech channel, we hypothesized that blind users can leverage concurrent speech to quickly get the gist of digital information. Grounded on literature reviews (e.g. [4,5,7,8]) that documented several features (e.g. spatial location, voice characteristics) that increase speech intelligibility, we investigated if and how we could take advantage of concurrent speech to accelerate blind people's scanning for digital information.

Results confirm that blind (and sighted [13]) people are able to scan for relevant content with two or three simultaneous voices [9]. Most importantly, we show [11] that two or three voices with speech rates slightly faster than the default rate, enable a significantly faster scanning for relevant content, while maintaining its comprehension. In contrast, to keep-up with concurrent speech timings, a single voice requires a speech rate so fast that it causes a considerable loss in performance. We then investigated and explored other prospective scenarios for concurrent speech interfaces. Besides scenarios that focus on information consumption, we explored the use of concurrent speech to support two-handed exploration in multitouch scenarios [10,12]. Overall, results show that concurrent speech is able to speed up the consumption of digital information in scanning scenarios, but that in tasks that require a greater physical coordination with the speech sources, the benefits are more dependent on the task itself and on user strategies.

Finally, we present a set of scenarios that emerged from a formative user study with blind participants and a set of open challenges in the use of concurrent speech to speed-up blind people's information consumption.

## Scanning for Relevant Content with Concurrent Speech

Sighted people's fast reading skills enable them to quickly get a general idea of the content – skimming – or to find specific information – scanning [1]. In our research, we refer to *Relevance Scanning* as the process of exploring the content and determine which pieces of information are relevant and deserve further attention.

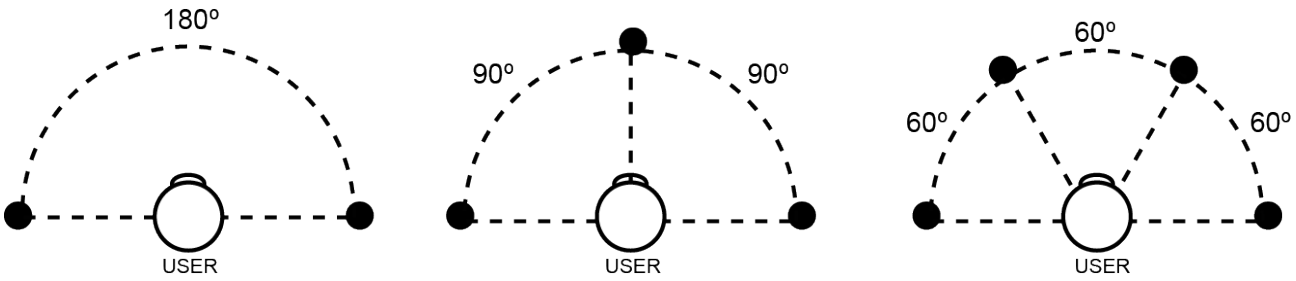


Figure 1. The spatial locations of the different speech sources used in the user studies reported in [9,11]

Literature reviews have documented the human ability to listen to concurrent speech and focus on a specific voice through selective attention. However, most of the previous experiments were performed with sighted users and used very short texts as input (e.g. from just using syllables to using 5-6 words). More often than not, digital information may comprise larger sentences whereas the conclusions of previous experiments cannot be applied. In our first study [9], we conducted an experiment aiming to understand blind people's ability to identify and understand relevant digital content listening to two, three and four concurrent speech channels (Figure 1). Results revealed that it is easy to identify one relevant sentence with two and (for most people with) three concurrent voices. Moreover, both two and three sources may be used to understand the relevant source content depending on the task intelligibility demands and user characteristics.

Considering the increasing popularity of auditory media among sighted people, we conducted a comparative analysis of blind and sighted people's perception of concurrent speech, where sighted people performed the exact same experiment previously performed by blind people [13]. Results support that both user groups are able to process concurrent speech in scanning scenarios. Moreover, the analysis showed no significant differences between groups, suggesting that sighted people may as well use two or three concurrent voices when there is the need to identify and/or understand the content of one relevant sentence. The absence of significant differences between these user groups promotes new approaches and interfaces that target wider audiences, rather than very specific solutions focused exclusively on blind people.

The ability to listen to two or three simultaneous sentences suggested that current screen readers may be imposing limitations on the way auditory feedback is being provided. While these results pointed out concurrent speech as a proper alternative to faster speech rates, only a direct comparison could determine their relative benefits and limitations. Therefore, we compared the use of concurrent speech against the use of faster speech rates when scanning for relevant digital information [11]. Moreover, we combined these two approaches by gradually increasing the speech rate with one, two, and three voices. Results showed that concurrent voices with speech rates slightly faster than the default rate, enable a significantly faster scanning for relevant content, while maintaining its comprehension. In contrast, to keep-up with concurrent speech timings, a single voice requires a speech rate so fast that it causes a considerable loss in performance. Overall, results suggest that the best compromise between efficiency and the ability to understand each sentence is the use of two voices with a rate of 1.75*default-rate (approximately 278 words per minute).

## Leveraging Two-Handed Exploration on Touchscreens

With a better understanding of how concurrent speech behaved in scanning scenarios, we started to investigate other contexts that could potentially benefit from the use of simultaneous audio

sources. In particular, we wanted to investigate this approach in scenarios that require more user interaction and control in order to receive the concurrent speech signals. A particular goal was to explore the usage of concurrent speech in multitouch interaction in touchscreen devices. While touchscreens support multitouch interaction, current screen readers are limited to a single, sequential auditory channel. However, the growing dimensions of touchscreen surfaces enables two-handed interaction and exploration of the screen. First, we supported two-handed exploration of large touch surfaces, using simultaneous, spatial audio feedback [12]. Then, we supported two-handed interaction in non-visual text entry on tablet devices through multitouch exploration and spatial, simultaneous audio feedback [10]. In these two scenarios we tried to understand and compare how blind people interact with one and two input (and feedback) points. Most importantly, we wanted to understand how concurrent speech could be used to support an additional input point in multitouch interaction. Results showed some advantages for the two-handed interaction regarding the ability to leverage the spatial knowledge of the screen. However, they have also shown that it is not trivial to coordinate the exploration of the screen with the simultaneous feedback, which may be explained by the high cognitive demands of both tasks. However, such demands seem to decrease when blind people use structured exploration strategies, which increases the benefits of concurrent speech.

## Prospective Scenarios for Concurrent Speech Interfaces

The first user studies revealed that the use of (faster) concurrent speech is able to speed-up the consumption of digital information while maintaining the basic understanding of the content. The results that posed concurrent speech as a strong alternative for Relevance Scanning scenarios, led us also to explore other interaction scenarios where it could be used to enhance blind users' digital experience. We built a user interface (The Cocktail Application) that supported Relevance Scanning in the contexts of news sites, Google search results, and e-mail. This application worked as a prompt to qualitative semi-structured interviews with 12 blind users to discuss and gather a set of scenarios that may take advantage of concurrent speech. Results revealed a tendency for Relevance Scanning scenarios when browsing lists of items, but they also exposed other scenarios where the use of a single auditory channel may be imposing limitations in the way users consume digital information. Herein, we describe potential scenarios mentioned by the participants of the user study, as well as other emergent scenarios brought out by discussions with other researchers in the field.

### Relevance Scanning

Listening carefully to documents, news, or blog posts require a person's attention and the use of concurrent speech would most likely hamper the full comprehension of the text. However, a preliminary selection task where users assess the worthiness of an information item does not require understanding the entire content. Among several news items, Google search results, e-mails, posts, links, or podcasts lies a decision of which are relevant and deserve further attention. This Relevance Scanning task is the scenario addressed in our first user studies and is currently done via the sequential audio of screen readers. Yet, the use of concurrent speech can accelerate this task in comparison to current solutions such as a single voice with faster speech rates [11].

The web accommodates a multitude of platforms that comprise numerous summarized, or already small per se, information items that (try to) provide the gist of the content to help deciding if they need further attention. These platforms may contain titles or small descriptions/snippets and include, for example, search engines, SNS such as Facebook and

Twitter, blogs, RSS feeds, news sites, and e-mail platforms. In this user study, besides the aforementioned applications and other specialized scenarios (e.g. shopping websites), participants' suggestions included site navigation through lists of links or headings. This scenario covers a multitude of websites and applications where blind users may navigate through menus or different sections, by listening to them simultaneously and selecting the one of interest. This user study suggests that in these scenarios users may process such lists for immediate consumption, to mark a set of relevant items for further analysis or even to get an overview of the content without acting on the data.

## Scanning for Specific Information

Websites, documents, or books with a lot of text may hinder the search for specific content when the user struggles to find a particular word or phrase to search for. Participants suggested the use of concurrent speech when scanning for a particular subject while studying or searching for something specific in a book. This may be especially useful in longer texts, where users may divide the content of different paragraphs, sections, chapters or even different websites into different concurrent voices. To cite a few examples, one could be searching for a particular detail on a Wikipedia page or an audio book, or open two websites from a Google search and start reading them to understand which one has the relevant information.

## Notifications using a Secondary Audio Channel

The aforementioned scenarios focus on scanning tasks that occur occasionally. The use of concurrent speech as the main mode to consume auditory information would be highly cognitively demanding and therefore somehow unrealistic. While the main exploration mode may still rely on a unique speech source, notifications do not need to be confined to uninformative alert sounds. While listening to a document, blog post or the daily news, chat or e-mail notifications could include the subject or the sender's name, instead of a beep sound that may induce the user to interrupt his current task. Moreover, when this information is provided, it should not interrupt what the user is doing (as stated by one participant when referring to Skype Talking). Instead, it could use a secondary channel to provide such notifications. Another example is the one of SNS, where a user may be listening to the news feed and simultaneously be informed about new notifications or chat alerts. Moreover, a proper use of Accessible Rich Internet Applications specification (WAI-ARIA) could leverage a secondary speech channel to help deal with dynamic content, website refreshes, and advanced interface functions developed with Ajax, HTML5, or JavaScript.

Although the use of a secondary channel to deal with notifications may benefit from the use of concurrent speech, it is important not to overload the user and distract (at least, too much) the user from the task at hand. That being said, there may be situations where users want to receive all notifications as they arrive. However, when performing tasks that require greater attention, notification management could gather and present a summary less frequently. On other occasions, notifications can be completely turned off. Another important aspect, as mentioned by one participant, is the type of notifications, since some may be so important that they should be read with full attention (e.g. system notifications).

## TV Navigation and Subtitles

The advances in digital television are excluding blind users from an equal access to the features they now provide to sighted users. While watching a particular channel, sighted users may navigate through the other channels, seeing what is playing on a particular channel without

actually changing the channel. However, this information is visual. Even after changing the channel, the information about the current (and following) program is still visual. The use of one Text-to-Speech channel would enable blind users to have the same feedback that sighted users have visually, while simultaneously presenting the regular audio of the current channel. Another scenario suggested was the one of subtitles. In this scenario, blind users could use headphones to hear both the regular sound of a movie/program and the subtitles when watching something in a language they do not understand.

Apart from these scenarios, participants also referred to the navigation on the TV and Video listings in the same way as in the Relevance Scanning scenarios previously described.

## Assisted Navigation

Participants referred to the GPS navigation while walking on the street as a good scenario for concurrent speech. In this case, the spatial location of the speech sources could be used to reflect the location of the items reported. However, it is important to notice that walking on the street may be a cognitively demanding task by itself. It would be interesting to understand how concurrent speech could be presented in this context without overloading the user. Actually, it would be interesting to understand how one (or two) voice could be combined with the environment sound, so that blind users may have a comprehensive auditory aid, without compromising the understanding of their surroundings. This may be achieved, for example, with bone-conducting headphones, which do not cover the ears and are also able to provide some auditory spatial cues [20] (although not as accurate).

## Text-Entry Correction and Feedback

Text entry in touchscreens is a highly demanding task. Besides being slow when inputting text in soft QWERTY keyboards (the defacto method), it is also error prone [18]. Alternative methods, such as braille-based keyboards (e.g. [19]) were able to improve input speed, but the typing accuracy remains a major problem. A common solution to deal with text entry inaccuracy is the use of spellcheckers, which flag words that may be spelled in an incorrect way and suggest alternatives. These correction systems are often based on features such keyboard layout and word frequencies. For example, B# is a correction system for multitouch Braille input that uses chords as the atomic unit of information rather than characters [17]. Although these correction systems already provide accurate suggestions that reduce the number of errors, research has not focused on how to present such suggestions to the user, including in mainstream solutions and methods. We are currently exploring [16] how to present suggestions in secondary auditory channels, while the main channel reads aloud the characters inserted.

## Collaborative Work

Tools such as Google Docs enable users to collaborate by editing the same documents either at different time periods or simultaneously. In these tools, sighted users are able to edit while seeing other users' activity, but this information is inaccessible to blind users. Although it seems unlikely that blind users would like to listen to everything as their co-workers write, concurrent speech could be used to provide more knowledge to the user about who's writing, when, and where in the document.

## Open Challenges

The findings of our research point towards the use of concurrent speech to speed up the auditory consumption of digital information. Herein, we present possible directions of future research that may extend this work or address topics that were elevated by our research.

**Study Interaction Mechanisms.** Our last user study enabled users to interact with concurrent speech in three different scenarios. However, users were only able to select and mark the items they were interested in analyzing with further care. Besides a deeper study of which commands are of most importance to deal with concurrent speech, there is the need to study how users can apply those commands in real world scenarios. The most prominent options are the ones of keyboard shortcuts and touchscreen gestures. However, within a multitude of commands that already exist, both in personal computers and mobile devices, the inclusion of new commands should be carefully studied to ease the navigation and interaction with multiple audio sources.

**Integration with Mainstream Screen Readers.** Participants' comments strengthened our stance that concurrent speech should be integrated in mainstream screen readers. Although focused, applicational solutions may benefit scanning in their particular scenarios, the impact of concurrent speech approaches can only be maximized if they are made available in the solutions that are transverse to their digital navigation.

**Study In-the-Wild Usage.** The integration with mainstream screen readers would ease studies that try to understand how people deal with concurrent speech approaches in real scenarios and when performing their typical interaction and navigation. In-the-Wild studies enable to capture usage and behaviors that are not possible to capture with laboratory-based evaluations [15]. Such studies can help understanding how and when concurrent speech is used and improve the users' experience based on their needs and interaction patterns.

**Explore Different Usage Scenarios.** In this research, we explored mostly *Relevance Scanning* scenarios, followed by two-handed interaction in touchscreen devices. In the previous section, we discuss several other scenarios that may benefit from the use of concurrent speech. Further research and solutions would help understanding which scenarios may also benefit from concurrent speech approaches.

**Study the Combination with Other Techniques.** This research showed that concurrent speech can be combined with faster speech rates and also with navigation through Headings. Such combinations can accelerate information scanning even more than each technique alone. Another research direction would be merging concurrent speech techniques with other promising solutions, such as summarization [1,14].

**Study the Effect of Practice and Learning.** There is evidence that speech segregation can benefit from practice [2]. In line with this evidence, our user study with faster concurrent speech suggested that participants were able to improve very slightly (non-significant) even with small speech rate increments. Since our user studies took approximately 45 minutes (on average), we were not able to assess the effect of practice on users' performance. However, such analysis

would help determining the *Information Bandwidth* that experienced users could reach and still maintain the basic understanding of the content.

**Study the Effect on Cognitive Load**. Participants' comments have suggested an increase in cognitive load when the number of voices (and speech rate) increases. Although some participants claimed they were a little tired at the end of the last trials, these experiments comprised very demanding conditions (particularly the last ones). Further research is needed, as their comments alone do not show the effect of concurrent speech over time in settings they find comfortable.

## Acknowledgments

The research presented above was performed during my PhD, always advised by Daniel Gonçalves. The user studies concerning two-handed exploration in multitouch scenarios were done in collaboration with Tiago Guerreiro, Hugo Nicolau, Kyle Montague, André Rodrigues and Rafael Nunes. This work was supported by the Portuguese Foundation for Science and Technology, under grants SFRH/BD/66550/2009 and INCENTIVO/EEI/LA0021/2014.

## References

1. Ahmed, F., Borodin, Y., Puzis, Y., & Ramakrishnan, I. V. (2012, April). Why read if you can skim: towards enabling faster screen reading. In Proceedings of the International Cross-Disciplinary Conference on Web Accessibility (p. 39). ACM.
2. Alain, C., Snyder, J. S., He, Y., & Reinke, K. S. (2007). Changes in auditory cortex parallel rapid perceptual learning. Cerebral Cortex, 17(5), 1074-1084.
3. Borodin, Y., Bigham, J. P., Dausch, G., & Ramakrishnan, I. V. (2010, April). 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) (p. 13). ACM.
4. Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. Acta Acustica united with Acustica, 86(1), 117-128.
5. Brungart, D. S., & Simpson, B. D. (2005). Optimizing the spatial configuration of a seven-talker speech display. ACM Transactions on Applied Perception (TAP), 2(4), 430-436.
6. Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. The Journal of the acoustical society of America, 25(5), 975-979.
7. Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. The Journal of the Acoustical Society of America,114(5), 2913-2922.
8. Ericson, M. A., Brungart, D. S., & Simpson, B. D. (2004). Factors that influence intelligibility in multitalker speech displays. The International Journal of Aviation Psychology, 14(3), 313-334.
9. Guerreiro, J., & Gonçalves, D. (2014, October). Text-to-speeches: evaluating the perception of concurrent speech by blind people. In Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility (pp. 169-176). ACM.
10. Guerreiro, J., Rodrigues, A., Montague, K., Guerreiro, T., Nicolau, H., & Gonçalves, D. (2015, April). TabLETS Get Physical: Non-Visual Text Entry on Tablet Devices. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 39-42). ACM.

11. Guerreiro, J., & Gonçalves, D. (2015, October). Faster Text-to-Speeches: Enhancing Blind People's Information Scanning with Faster Concurrent Speech. In Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (pp. 3-11). ACM.
12. Guerreiro, T., Montague, K., Guerreiro, J., Nunes, R., Nicolau, H., & Gonçalves, D. J. (2015, November). Blind People Interacting with Large Touch Surfaces: Strategies for One-handed and Two-handed Exploration. In Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces (pp. 25-34). ACM.
13. Guerreiro, J., & Gonçalves, D. (2016). Scanning for Digital Content: How Blind and Sighted People Perceive Concurrent Speech. ACM Transactions on Accessible Computing (TACCESS), 8(1), 2.
14. Harper, S., & Patel, N. (2005, October). Gist summaries for visually impaired surfers. In Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility (pp. 90-97). ACM.
15. Montague, K., Nicolau, H., & Hanson, V. L. (2014, October). Motor-impaired touchscreen interactions in the wild. In Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility (pp. 123-130). ACM.
16. Montague, K., Guerreiro, J., Nicolau, H., Guerreiro, T., Rodrigues, A., & Gonçalves, D. (2016). Towards Inviscid Text-Entry for Blind People through Non-Visual Word Prediction Interfaces. CHI Workshop on Inviscid Text-Entry and Beyond.
17. Nicolau, H., Montague, K., Guerreiro, T., Guerreiro, J., & Hanson, V. L. (2014, April). B#: chord-based correction for multitouch braille input. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1705-1708). ACM.
18. Oliveira, J., Guerreiro, T., Nicolau, H., Jorge, J., & Gonçalves, D. (2011, October). Blind people and mobile touch-based text-entry: acknowledging the need for different flavors. In The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility (pp. 179-186). ACM.
19. Southern, C., Clawson, J., Frey, B., Abowd, G., & Romero, M. (2012, September). An evaluation of BrailleTouch: mobile touchscreen text entry for the visually impaired. In Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services (pp. 317-326). ACM.
20. Walker, B. N., Stanley, R. M., Iyer, N., Simpson, B. D., & Brungart, D. S. (2005, September). Evaluation of bone-conduction headsets for use in multitalker communication environments. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 49, No. 17, pp. 1615-1619). SAGE Publications.

## About the Authors:

João Guerreiro received his Ph.D. in Information Systems & Computer Engineering from the University of Lisbon, advised by Daniel Gonçalves. His PhD focused on the use of concurrent speech to speed-up blind people's scanning for relevant digital content. He is currently a post-doctoral researcher in the Intelligent Agents and Synthetic Characters Group (GAIPS), INESC-ID Lisboa.