**The Use of Concurrent Speech to Enhance Blind People’s Scanning for Relevant Information**

João Guerreiro  
INESC-ID / Instituto Superior Técnico, Universidade de Lisboa  
joao.p.guerreiro@ist.utl.pt

**Abstract**

Screen readers present the information sequentially, which contrast with the visual presentation on screen. This fact affects blind users’ efficiency browsing digital information, particularly when performing fast reading tasks. In this thesis, we will study the use of concurrent speech sources to help blind people finding their information of interest faster. Our first results suggested that blind users are able to identify and understand a relevant speech signal when listening to 2 or 3 concurrent talkers. Our next steps include comparing with current alternatives; assessing the learning effect with practice; and finding ways to interact with those sources.

**Introduction**

Great efforts have been made to enhance the access to digital information by visually impaired users. Moreover, such users develop their own tactics when tackling problematic situations [14] and develop browsing strategies to scan the information more efficiently (e.g. navigate through headings or increase the speech rate) [4]. However, “the biggest problem in non-visual browsing remains the speed of information processing” [4], which results in significant differences in prejudice of blind users when compared to sighted users browsing performance [3]. For instance, webpages such as news sites and Social Networking Sites present several summarized information items, where the user needs to navigate in order to assess which are relevant and deserve further attention [10].

The differences between sighted and blind users are influenced by the fast reading techniques that sighted users employ, such as looking through titles and other visually prominent content or catching separate phrases (the so called diagonal reading) [1]. These techniques enable sighted users to sift through a website quickly to get an overview of its content (skimming) or to find specific information (scanning). In contrast, screen readers present the information sequentially to blind users, which contrast with the visual presentation on screen that portrays much more information at a time.

We believe that using multiple simultaneous speech sources is the pathway to enhance blind people’s ability to identify, quickly, their information of interest. In a recent study [11], with 23 participants, we aimed at understanding blind people’s ability to search for relevant news snippets while listening to 2, 3 or 4 concurrent speech channels. The results suggested that the identification of one relevant source is a straightforward task whilst listening to 2 simultaneous talkers and most participants were still able to identify it with 3 talkers. Moreover, both 2 and 3 simultaneous sources may be used to understand the relevant source’s content depending on speech intelligibility demands (how much needs to be perceived) and user characteristics.

In our next steps, we will compare our approach with currently used techniques, in particular with the use of fast speech rates. Finally, we will explore different ways to interact with these simultaneous sources.
Background
The *Cocktail Party Effect* states the human ability to focus the attention on a single person among several conversations and background noise [8]. Moreover, one may detect interesting content in the background (e.g. own name or favourite subject) and shift the attention to another person. In fact, several studies about the ear physiology and neuroscience address the human auditory system ability to segregate speech [5] and support the use of simultaneous speech sources. Yet, the speech sources configuration is crucial to maximize the number of concurrent talkers. Attributes such as the spatial location [6] and the voice characteristics [9] of each source play an important role in the segregation of a mixture of sounds into different streams.

To sum up a large number of previous studies, most of them focus on selective attention tasks, where participants have to report speech of a specific talker, neglecting the others. These studies (and even those on divided attention – e.g. [2]) usually require the exact identification of small phrases. In contrast, we believe that bigger texts will provide the gist of each source without listening to the entire content, in order to assess its relevance (*Relevance Scanning*) or to find specific information (*scanning*). What is more, most of the studies resorted to sighted users, but there is evidence that blind people’s sensory compensation enhances their ability to segregate, understand and report speech [12]. This fact is clearer for congenitally/early-blind people due to the process of *Neuro-Plasticity* that occurs until 18-21 years old [7].

Although there is valuable research that considered these studies insights (e.g. [13]), they end up neglecting the speech sources parameterization and focus on rather different contexts than scanning.

Goals and Contributions
In this research, we plan to investigate the opportunities to maximize the audio channel to provide more information to blind users, in order to allow them to find their information of interest faster. We believe concurrent speech will empower faster insights of the available information and a prompt decision about the ones of interest. Our hypothesis is: **Delivering multiple simultaneous parameterized sound sources to blind people enables faster but still effective scanning to find relevant information.**

**Concurrent Speech in Scanning Scenarios**
The use of simultaneous speech sources may not fit every scenario. We believe this approach will be more appropriate for *scanning* rather than for *skimming* tasks, since the latter requires an overall understanding of the texts. *Scanning* tasks may refer to finding specific information that a person is already searching for or finding information of interest without previous clues.

In our recent study [11], 23 participants had to listen to 2, 3 or 4 concurrent speech sources (news snippets) and identify and understand a specific one of them. A hint priming the relevant snippet was given before each trial. The speech sources spatial locations took inspiration from several experiments that use equally spaced positions in the frontal horizontal plane. Although other spatial configurations were proposed and provided better results overall, they ended up sacrificing specific locations [6]. In our spatial setting, the sound sources were separated by 180º, 90º and 60º, for 2, 3 and 4 talkers, respectively (Figure 1).

Although most experiments on simultaneous speech segregation focus on pitch variations, the best results are achieved when varying the two main characteristics that influence male and female voices - the pitch (Glottal Pulse Rate) and formant frequencies (Vocal Tract Length) [9].
We wanted to validate if these differences are also valuable for longer speech signals. Thus, we selected three different configurations: all speech sources with the same voice; larger pitch and formant frequencies separation; smaller pitch and formant frequencies separation.

Our results showed that identification of the relevant source is a straightforward task when listening to 2 talkers, and for most participants, it was also easy to identify with 3. Moreover, both 2 and 3 simultaneous sources may be used to understand the relevant source content, but depend on how much information needs to be understood and on user characteristics (working memory and attention). Unlike the related work, differences in voice characteristics did not provide a greater effect in neither speech identification nor intelligibility. However, participants preferred the use of concurrent talkers with different voices.

Sighted, Early and Late Blind
Our previous study did not present statistical differences between early and late blind participants. We are currently performing the same study with sighted people, which may point towards the use of simultaneous sound sources for such population in eyes-free interaction. This last comparison targets the frequent need to consume information, even when it is not possible to look at the screen/paper.

Comparison with Faster Speech Rates
We will compare the use of concurrent speech sources with faster speech rates. We will try to understand both their effectiveness and their effect on cognitive overload and fatigue. Moreover, we will include hybrid solutions that combine both. Finally, we will investigate the effect of learning and experience in the use of concurrent speech in scanning scenarios.

Interacting with Concurrent Speech
Our first study guided us towards the use of multiple sound sources. In line with these results, the interaction with one sequential information channel may be replaced with an interaction with (2 or 3) concurrent channels in fast reading scenarios. Current solutions rely mostly on head-tracking or pointing in order to select the current focus. We will investigate the best ways to control concurrent speech, which may include selecting or pausing a particular source, reducing/increasing the number of sources or even controlling their volumes.

Conclusions
At this stage of my research, it was very important to obtain feedback from accessibility experts and peers. This doctoral consortium helped me validating and improving our future research methodology. Our first results suggested that blind users are able to identify and understand a relevant speech signal when listening to 2 or 3 concurrent voices. In future work, we will continue to explore the perception of concurrent speech and we will try to find ways to interact with simultaneous sources.
**Acknowledgments**

We thank the Fundação Raquel e Martin Sain, Carlos Bastardo and all the participants of our studies. This work was supported by national funds: FCT- Fundação para a Ciência e Tecnologia, under individual grant SFRH/BD/66550/2009.

**References**


**About the Authors:**

João Guerreiro is a PhD Student at Instituto Superior Técnico (IST), University of Lisbon (UL). He is also a research assistant at the Visualization and Intelligent Multimodal Interfaces (VIMMI) group from INESC-ID. His main research interests include human-computer interaction and accessibility. In particular, he is currently exploring the use of concurrent speech to enhance blind people’s efficiency when scanning for relevant digital content.