

# COMPRESSING INFORMATION DENSITY IN AUDIO-VISUAL SENSORY SUBSTITUTION OF BLIND INDIVIDUALS

*Eric Heng Gu<sup>1</sup> and Charles Leclercq<sup>2</sup>*

<sup>1</sup>TU Delft, Netherlands

<sup>2</sup>ARx, UK

`h.gu@tudelft.nl, charles.leclercq@network.rca.ac.uk`

## ABSTRACT

In-vehicle intelligent voice agents can be beneficial to drivers as a companion in foreseeable autonomous driving. Speech style and From the invention of Braille in 1824[1] to the first wearable tactile interface in 1964 by Bach-y-Rita[2], sensory substitution techniques and devices have been explored to help blind individuals. On top of the learning curve of these novel interfaces, our processing capacity for visual information is magnitudes greater than auditory[3] or haptic stimuli[4], and to a higher degree of accuracy; We can at a glance read a sentence that would take several seconds to listen to. For blind individuals, transformation of visual data into a form that they can decipher becomes crucial to supporting their daily lives. Recent improvements in Natural Language Processing (NLP) offers us a new approach to compressing the informational density. The present work is designed for time savings when blind individuals interpreting a large selection of textual data, such as a newspaper or email inbox, by leveraging a fine-tuned Transformer model to condense the text into a short summary that can be converted to audio. We put forward a demonstration using the ARX Vision wearable headset[5], with YOLO[6] and OpenCV[7] detection modules.

