# SlAide2Voice: a new educational tool for students with visual disabilities

Giorgio Ciano[a,b], Giovanna Maria Dimitri[a], Alberto Rossi[a,b], Giorgia Giacomini[c], Simone Bonechi[a], Paolo Andreini[a] and Elisa Messori[a]

[a]University of Siena, Dept. of Information Engineering and Mathematics, Via Roma 56, Siena, Italy
[b]University of Florence, Dept. of Information Engineering, Via di Santa Marta, Florence, Italy
[c]University of Siena, Dept. of Biotechnology, Chemistry and Pharmacy, Via Aldo Moro 2, Siena, Italy

**Abstract**

Online lessons had become more and more frequent and this new way of teaching, forced by the Covid-19 epidemic, implies many problems for all of the students and above all for those ones affected by disabilities. It is therefore absolutely necessary to address this issue and provide practical solutions for it. This paper proposes SlAIde2Voice, a new software pipeline architecture to help visually impaired students to overcome some difficulties related to online lessons. Our aim is to develop a tool, based on three simple components, a client–server architecture, an OpenOffice add-on and an artificial intelligence module, which will be able to help visually impaired students not only during online lessons, but also during their independent study. The proposed methodology promises to improve the learning quality of students with visual difficulties and aims to improve their inclusion and independence. SlAIde2Voice will be designed to be used together with any existing videoconference tools and its use can also be extended to conferences and meetings in general.

**Keywords**

Distance Learning, Blindness aid, Artificial Intelligence

## 1. Introduction

### 1.1. Motivation

Due to the Covid-19 outbreak, the role played by technology in teaching and learning activities is becoming more and more predominant. School closures has required the implementation of distance learning solutions. All around the world, millions of students experimented for the first time with this new way of attending school. This way of participating lessons involves not only students, but also their families, providing extra difficulties for some types of disabilities, such as visual impairments.
The impact of COVID-19 on students with disabilities could be manifold. In fact such students might suffer due to the lack of accessible software, teaching materials and tutor direct support [1]. The World Health Organization (WHO) has estimated that globally 285 million people of

*First Workshop of teleXbe, Technology Enhanced Learning Environments for Blended Education, University of Foggia and Smarted s.r.l., 21-22 January 2021*
giorgio.ciano@unifi.it (G. Ciano); giovanna.dimitri@unisi.it (G. M. Dimitri); alberto.rossi@unifi.it (A. Rossi); giacomini@student.unisi.it (G. Giacomini); simone.bonechi@unisi.it (S. Bonechi); paolo.andreini@unisi.it (P. Andreini); elisa.messori97@gmail.com (E. Messori)

all ages are visually impaired, of which 39 million are blind. Furthermore, almost 18.9 million children below the age of 16 have visual disabilities[2]. Generally, blind or visually impaired students need teaching material presented through other channel by means of a tactile sign language interpreter to facilitate communication and learning in the school environment.
This paper address the problem of replace visual channel in remote lesson, since vision is fundamental to fully comprehend online presentation, even more considering that a tutor could not be available due to covid restrictions. According to [3], a large portion of students feels Sars-Cov-2 influenced the decision to continue their educational path (11% students decided to not enroll in college and 24% of students claimed they were likely to change their minds about what college to attend) and the majority of college students (63%) think online classes are less effective with respect to traditional in-person lectures. Moreover, parents pointed out the difficulty to reach the teaching staff and to be assisted by it. Finally, it is important to emphasise that distance education courses have been largely used also in pre-pandemic years and may increase in the next years [3]. It is clear, therefore, that allowing visually impaired students to have the same material of the rest of the class is necessary, especially when dealing with distance learning.
This is not a problem that affects only people at young age, since videoconferencing programs, and remote presentations, are largely used also in work environments.
The ability of access completely the material provided by the teacher, or by colleagues in a company, could be game-changing for people around the world who are visually impaired, counting both blindness and low vision.

### 1.2. Supports for learning targeted at visually impaired people

Listening to the teacher's voice may not be enough for visually impaired students, especially in the case of remote teaching, where audio quality could be compromised due to connection problems or sound overlapping (e.g. more students talking at once). Some tools are already available to assist visually impaired and blind students [4]:

- Text-to-speech software are programs capable to convert textual information into audio, and provide therefore material to the visually impaired users. Such tools may be not particularly effective when practicing a foreign language, but they would allow the user to operate on a braille display simultaneously.
- Braille display,is a computer device that converts in braille the text on the screen. It usually consists of a line of cells with eight holes, from which small cylinders can come up to form symbols of the braille alphabet. In order to properly work, the braille display needs to use a screen reader that selects on the screen the information to be transformed into braille time by time.
- Screen reader, provides a speech synthesizer that can describe to the user menus, buttons and in general tools to move inside of an application or website.
- Braille printers, able to print braille characters and figures embossed, using the six points code.

These tools effectively help visually impaired students to use the same teaching material prepared for the whole class by the teacher as long as the materials can be converted into digital

form. However, this kind of materials frequently contains images and graphs which are difficult to be translated using the mentioned tools. Indeed, as said in [4], texts integrated by a large use of images still constitutes an obstacle for the learning process of the unsighted student.
In addition, the tools available can help students while studying but cannot be effectively used for online lessons, while having an online conversion to braille via a braille tablet during the lesson could be a vital help for students. For this reason SlAide2Voice include an artificial intelligence module to extract the description from the images and the graphs and translating all the information contained in the slides, in Braille, during the lesson or presentation, strengthen the integration of visually impaired students throughout the course.
It is worth noting that among the tools that offer assistance to blind students or people with vision deficiency, there is also Blackboard Ally, a software that focuses on making digital content more accessible. However, this model aims at facilitating the exploitation of the teaching material interfacing with Moodle only in offline way: on the other hand, SlAide2Voice has the goal of helping the user (a student or anyone in a videoconference meeting) during the lecture or meeting, interfacing with any videoconferencing program. SlAIde2Voice leverages Artificial Intelligence (AI) technologies to provide educational material to visually impaired students in an easily accessible way. In recent years, AI and deep–learning [5] in particular have made extraordinary progress. This has allowed the diffusion of these technologies in a wide variety of domains, ranging from the development of support systems for medical diagnosis [6, 7], the analysis of graphs [8, 9, 10, 11] to the development of generation systems able to synthesize a multitude of data like images and sounds [12, 13, 14]. Advanced artificial intelligence systems can make it possible to transform images and written text into spoken language. SlAIde2Voice leverages these technologies to provide educational material to visually impaired students in an easily accessible way. The paper is organized as follows. The proposed architecture together with its submodules has been presented in Section 2, while, in Section 3, the impact that SlAIde2Voice could have on visually impaired students was discussed. Finally, conclusions and future researches have been collected in Section 4.

## 2. Materials and Methods

### 2.1. OpenOffice

OpenOffice is a free productivity suite developed by the Apache Software Foundation and is one of the best alternatives to Microsoft Office. This suite is available for various operating systems, including Apple MacOS, Microsoft Windows and Linux. Within this software are present different types of applications, all very intuitive and easy to use:

- WRITER, for word processing or desktop publishing.
- CALC, for developing spreadsheets.
- IMPRESS, to create effective presentations in a simple way.
- DRAW which offers tools for creating graphs and diagrams.
- BASE for database management.
- MATH for mathematical equations, very useful as equation editor for text documents.

There are many advantages to using OpenOffice, and open source software in general. First of all, there is a cost saving, since it is completely free software, and this allows us to use that saved money to buy hardware useful for our purpose. Secondly, the Apache 2.0 License allows you to use the software for any purpose, including commercial, and you can install it on as many computers as you like. Furthermore, with this license you can freely customize the open-source code, so it allows you to modify and maybe even improve the software and make it available to the public. For our purpose, the most important feature of this suite is the opportunity to create extensions and then add functions to the existing software.

### 2.2. Braille Tablet

Braille code is one of the main tools used by blind people to access written information. In the recent years, there has been a technological effort to develop refreshable Braille devices. These devices consist of multiple physical dots that dynamically change their configuration to reproduce different sequences of the letters in Braille code. A few multi-cell/multi-line devices can be found in the literature. For example one of the most common refreshable braille disposals are based on pixo-electric actuators. Such types of implementation show good displacements of the pins and high rate of refreshing the raising dots, but due to their technological nature they usually require big cell dimensions and high operating voltage, resulting in high costs [15][16]. Instead, the use of shape memory alloy (SMA) coils has been proposed in [16]. Unfortunately it has been shown to exhibit low portability, and negative aspects due to their high working temperatures, which could imply several cooling procedures to be used. Other solutions focus on the use of a flapper mechanism for each dot of the Braille cell [17]. Many approaches have been proposed for multi-cell devices while single-cell refreshable systems has received little attention so far. In [18] a new cost-effective single-cell Electromagnetic Refreshable Braille Display has been proposed. While single–cell devices are already present in the literature [19], the solution proposed by [18] is more cost-effective. Moreover experimental evidence proved the efficacy of the proposed system [18] for reproducing alphanumeric content.

### 2.3. Image Captioning and Text–to–Speech

Image Captioning is the process of generating natural language descriptions of an image. Image captioning requires to recognize the important objects, their attributes, and their relationships in the image. The information extracted from the image have then to be converted in natural language. The generated sentences needs to be syntactically and semantically correct. For this reason, the development of effective image captioning systems is considerably challenging because it combines computer vision and natural language. The recent introduction of deep–learning–based techniques allows to handle these complexities and challenges. State–of–the–art deep–learning–models for image captioning are approaching the human level performances in this task [20]. A comprehensive review of image captioning techniques is beyond the scope of this work but for an interested reader a detailed survey can be found in [21]. Image captioning can be integrated with text–to–speech technologies to produce a ”spoken description” of an image. Text–to–speech consists in synthesizing artificial human speech from text. It is an essential component in many applications that provide accessibility for the visually–impaired because it allows human–technology interaction without requiring visual interfaces. Despite decades of investigation, generating high quality natural speech from text, remains a challenging task but the introduction of deep–learning has been revolutionary also in this field allowing to generate very high-quality text–to–speech [22]. Some commercial tools exploit these recent advances to provide support for the visual impaired. One of the most prominent example is the *seeing AI*, an artificial intelligence application developed by Microsoft. *Seeing AI* uses the device camera to identify people and objects, and then the app audibly describes those objects for people with visual impairment.

## 2.4. Proposed Architecture

The architecture we aim to develop can be used in conjunction with any existing videoconferencing tool on the market to help people with visual impairments during online lessons, conferences and so on. The three main components of the software are the following:

- A simple client-server interface, which allows the exchange of data between the presenter and the students.
- An OpenOffice add-on, which has to capture the slide change and extract its content (text and images).
- An AI model capable of generating captions and descriptions of the images contained in the slide.

The software must be installed on both the computer of the speaker and of the blind students. The presenter must upload the slides before the lesson to be pre-processed by the application. In this phase, the software generates the metadata, in particular, using the OpenOffice add-on, it extracts text, images, and the layout of each slide. Next, a deep convolutional neural network will be used to generate the image caption and extract its description. Image borders will be also collected to convert the images into a format that can be represented on the braille tablet. Finally, all metadata will be embedded in an XML file that the software shares with students at the beginning of the lesson. The main functionalities are also shown in Figure 1, and in particular, the system could work in two main ways:

- **Online** – During the video call, it automatically translates the current slide, shown on the screen, into braille using the tablet. In this way, visually impaired people were able to better contextualize the concepts expressed by the speaker.
- **Offline** – At the end of the call, the software automatically generates a package containing the slides and metadata (e.g. audio of the lesson and description of the slides). In this way, we will give students a platform that can be used at home, while studying, allowing them to be more independent.

In particular, as regards the ”Online” mode, once the slide and the metadata have been shared, the software will synchronize the presented slide between the speaker and the students, through the client–server environment. Then, from the student’s side, it sends the metadata, converted to braille, to the tablet. At the end of the lesson, in ”Offline” mode, the software automatically matches the audio of the lesson to the slide, creating an ”all in one” package containing all the

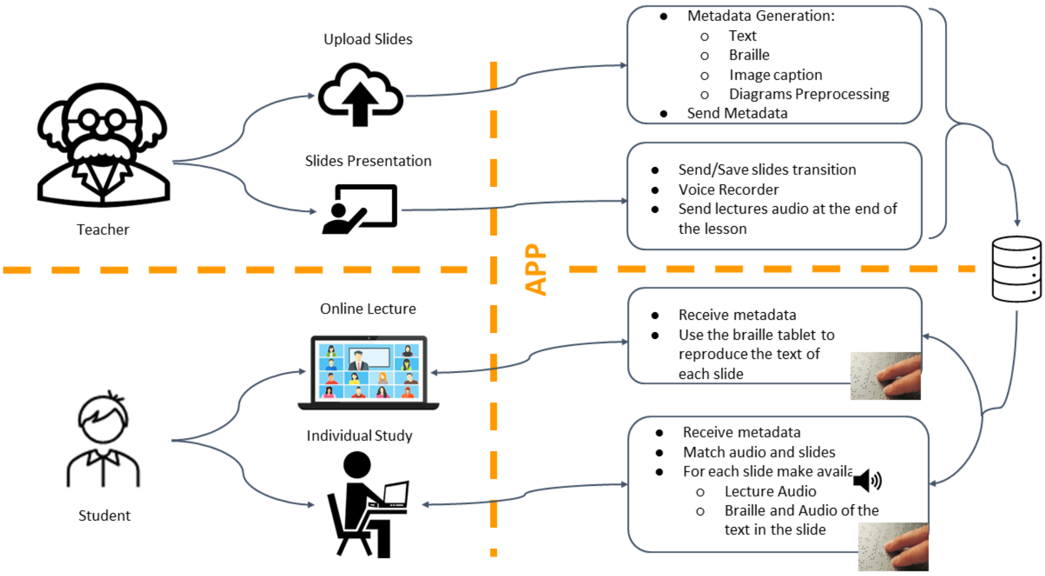


**Figure 1:** SlAIde2Voice main functionalities

material of the lesson. This package will be shared with the students and, once opened with our application, the student will be able to scroll the slides and decide to listen to the teacher's explanation and to reproduce the slide content on the braille tablet or with a synthesized voice.

## 3. Discussions: strengths and limitations

### 3.1. The impact

According to the report proposed by the world health organization[23], people affected by visual impairments were estimated to be 285 millions in 2010, 39 millions of whom were completely blind. Incidence of visual impairments in the younger population is also very high, with an estimate of approximately 19 millions of people in the age range $0-14$ of whom 1.5 millions completely blind [23]. Significant costs are related to the development of education tools dedicated to visually impaired people. In [24] the authors report a total amount of 615 millions of dollars spent in education devices for children in age range $0-17$. This segment of the paediatric population is composed by approximately 3000 children, and the estimate is that each individual will spend around 0.2 millions of dollars in the 17 years. We believe therefore that our system SlAIde2Voice, based on braille board and the related software, could dramatically improve the quality of the education received by a visual impaired student, together with a massive expenses reduction. Using our tool, students can easily translate slides into vocal learning material or directly to the braille board. In this way students can simultaneously listen to the presenter, and have the haptic "written" braille device which could replace the text

material present in the slides.
Assuming a cost of around 500 dollars for the braille boards, we can already grasp how the costs of the system proposed would be low, and therefore the system easily accessible to a wide section of the population.

### 3.2. Open-source based platform

SlAide2Voice architecture, will be entirely based on open-source platform. This is a huge strength, since it allows to have no need of specific licenses and it therefore reduce dramatically the costs related to those. Moreover it is not related to a specific slide maker software. This is also a very important point to be considered. In this way, in fact, no agreements are required with a specific software house, and this aspect allows to have no issues concerning administrative and legal implication of using non open-source programs. Moreover this can also efficiently reduce the time of production phase for the system itself, in a very effective way.

### 3.3. Cross-platform

Another key point of strenght of Slaide2Voice is the capability of working closs-platform, with any conference system used for the online lesson. Nowadays, in fact, a lot of different systems for web meeting are used in different contexts, and therefore there is the need of developing something which takes into account such aspect. To tackle this, we decided to follow the strategy of ”by-passing” the conference system platform, using a modular approach for the software architecture implementation. Each functionality of the system will be embedded in a separate module, which could be integrated in open-office, or implemented by scratch during the development process of SlAide2Voice, which will therefore become free from integration issues with proprietary software and specific architectures.

### 3.4. Braille Devices

Surrogating visual contents for blind and visually impaired people, through Braille scripting, is a key aspect of the software SlAide2Voice. This will allow the user to have an immediate live haptic feedback of the visual content which is shown on the slide currently projected by the presenter. The choice of representing Braille code through haptic devices has been shown to be successful not only for text information, but also to translate visual (graph and figures) as well as virtual objects representations [25][26]. When used in a live environment, such as during a conference or online presentation, refreshable Braille displays represent a good technical solution to obtain the high performances. In refreshable Braille devices the typical 3x2 or 4x2 Brailles dotted cells are represented through dynamically raised dots [18]. In this way the SlAide2Voice user could listen to the voice of the presenter and simultaneously have the availability of the Braille text which is projected in the slide, including figures and tables. In this way comprehension could be enhanced. As several researches in the field show, in fact, the combination of haptic and audio stimuli improve significantly the understanding of the content explained [27][28]. We considered several option as candidate to be integrated in SlAide2Voice. As first option we considered the use of multi-cell/multi-line devices.

Unfortunately one of the main drawbacks of these systems is the high cost of production, which could represent a dramatic drawback for the implementation and integration in SlAide2Voice. This is the reason why we looked for possible alternatives to be integrated in the system. One solution which could dramatically reduce costs is the single cell technology. In particular it can reduce the costs related to design complexity, portability and power usage [18]. An interesting solution in this sense is the device ”Readable”, described in [18].

Indeed, Braille devices usually range between 500$ and 2000$, while ”Readable” has an overall cost of 42.74$ [18], which the authors claim to be comparable to cost-aware devices.

The extension of Readable would allow to reproduce also graphic contents in Braille, as needed by our software SlAide2Voice.

### 3.5. 3D printing

For what concerns the offline translation of slides into braille text, an alternative could come from 3D printing, nowadays extremely popular for modelling and prototyping. For example in [29] the authors present a computer-based approach for 3D printing accessible to blind People to explore systems which would increase the self independence of the users. Cost and need of developing expertise for the use of the 3D printing software, represent however a drawback to be taken into consideration when evaluating 3D printing as a possible option to be included in SlAide2Voice [30].

## 4. Conclusions

We proposed a new framework to improve the fruition level of online lecture for visually impaired students. Thanks to precise design choices, our tool is completely independent from the conference platform, making the environment extendable to every person affected by visual disabilities who would like to join an online meeting. Moreover it allows to reproduce slides in an offline way, something particularly useful for self-study or reasoning after the presentation. Finally the use of open-source slide making software avoid the problems related to licensed suite, indeed reducing the realization costs to the braille tablet purchase only. We believe that SlAide2Voice could fill the gap in learning tools for visually impaired people, especially in conditions in which real lectures and conferences are not possible, as in the case of Covid-19. Our proposed software can increase the self-independence of the user and therefore reduce the need of help from parents or tutors. Moreover, thanks to the acquaintance of self-independence, national money devoted to instantiate special dispositions for visually impaired patients can be saved. Costs, could be even more reduced with the availability of cheaper braille tablets in the market. We think that our system could be very useful, also when Covid-19 pandemic will be over, for normal lecturing or already existing online university courses (i.e Coursera, Udemy[1]). The approach used to develop SlAide2Voice architecture, can easily be extended to support learning with other disabilities. For instance, the integration of a simple speech recognition system make the software useful for hearing impaired people.

[1] www.coursera.org www.udemy.com

## References


[1] W. H. Organization, Policy brief: The impact of covid-19 on children (2020). URL: https://unsdg.un.org/sites/default/files/2020-05/Policy-Brief-A-Disability-Inclusive-Response-to-COVID-19.pdf.

[2] W. H. Organization, Blindness and vision impairment (2020). URL: https://www.who.int/en/news-room/fact-sheets/detail/blindness-and-visual-impairment.

[3] J. Bustamante, Online education statistics (2020). URL: https://educationdata.org/online-education-statistics.

[4] U. C. d. S. C. Michelle Pieri, Fondazione Don Carlo Gnocchi, Tecnologie per l'inclusione scolastica degli alunni con disabilità visiva (2020). URL: http://portale.siva.it/files/doc/library/a423_1_A12_Pieri_Michelle.pdf.

[5] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (2015) 436–444.

[6] Y. Li, L. Shen, Skin lesion analysis towards melanoma detection using deep learning network, Sensors 18 (2018) 556.

[7] P. Andreini, S. Bonechi, M. Bianchini, A. Mecocci, F. Scarselli, Image generation by gan and style transfer for agar plate image segmentation, Computer Methods and Programs in Biomedicine 184 (2020) 105268.

[8] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, G. Monfardini, The graph neural network model, IEEE Transactions on Neural Networks 20 (2009) 61–80.

[9] N. Pancino, A. Rossi, G. Ciano, G. Giacomini, S. Bonechi, P. Andreini, F. Scarselli, M. Bianchini, P. Bongini, Graph neural networks for the prediction of protein–protein interfaces, ESANN 2020 proceedings (2020).

[10] T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: International Conference on Learning Representations (ICLR), 2017.

[11] A. Rossi, M. Tiezzi, G. M. Dimitri, M. Bianchini, M. Maggini, F. Scarselli, Inductive–transductive learning with graph neural networks, in: IAPR Workshop on Artificial Neural Networks in Pattern Recognition, Springer, 2018, pp. 201–212.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in neural information processing systems, 2014, pp. 2672–2680.

[13] P. Andreini, S. Bonechi, M. Bianchini, A. Mecocci, F. Scarselli, A. Sodi, A two stage gan for high resolution retinal image generation and segmentation, arXiv preprint arXiv:1907.12296 (2019).

[14] L.-C. Yang, S.-Y. Chou, Y.-H. Yang, Midinet: A convolutional generative adversarial network for symbolic–domain music generation, arXiv preprint arXiv:1703.10847 (2017).

[15] R. Sarkar, S. Das, Analysis of different braille devices for implementing a cost-effective and portable braille system for the visually impaired people, International Journal of Computer Applications 60 (2012).

[16] D. Leonardis, L. Claudio, A. Frisoli, A survey on innovative refreshable braille display technologies, in: International Conference on Applied Human Factors and Ergonomics, Springer, 2017, pp. 488–498.

[17] C. Loconsole, D. Leonardis, M. Gabardi, A. Frisoli, Braillecursor: An innovative refreshable braille display based on a single sliding actuator and simple passive pins, in: 2019 IEEE

World Haptics Conference (WHC), IEEE, 2019, pp. 139–144.
[18] G. C. Bettelani, G. Averta, M. G. Catalano, B. Leporini, M. Bianchi, Design and validation of the readable device: A single-cell electromagnetic refreshable braille display, IEEE Transactions on Haptics 13 (2020) 239–245.
[19] N. Sojib, M. Z. Iqbal, Single cell bangla braille book reader for visually impaired people, in: 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), IEEE, 2018, pp. 1–4.
[20] X. Hu, X. Yin, K. Lin, L. Wang, L. Zhang, J. Gao, Z. Liu, Vivo: Surpassing human performance in novel object captioning with visual vocabulary pre-training, arXiv e-prints (2020).
[21] M. Z. Hossain, F. Sohel, M. F. Shiratuddin, H. Laga, A comprehensive survey of deep learning for image captioning, ACM Comput. Surv. 51 (2019). URL: https://doi.org/10.1145/3295748. doi:10.1145/3295748.
[22] S. O. Arik, M. Chrzanowski, A. Coates, G. Diamos, A. Gibiansky, Y. Kang, X. Li, J. Miller, A. Ng, J. Raiman, et al., Deep voice: Real-time neural text-to-speech, arXiv preprint arXiv:1702.07825 (2017).
[23] D. Pascolini, S. P. Mariotti, Global estimates of visual impairment: 2010, British Journal of Ophthalmology 96 (2012) 614–618.
[24] J. S. Wittenborn, X. Zhang, C. W. Feagan, W. L. Crouse, S. Shrestha, A. R. Kemper, T. J. Hoerger, J. B. Saaddine, V. C.-E. S. Group, et al., The economic burden of vision loss and eye disorders among the united states population younger than 40 years, Ophthalmology 120 (2013) 1728–1735.
[25] J. J. Zárate, H. Shea, Using pot-magnets to enable stable and scalable electromagnetic tactile displays, IEEE transactions on haptics 10 (2016) 106–112.
[26] L. Brayda, C. Campus, M. Memeo, L. Lucagrossi, The importance of visual experience, gender, and emotion in the assessment of an assistive tactile mouse, IEEE transactions on haptics 8 (2015) 279–286.
[27] A. Russomanno, S. O'Modhrain, R. B. Gillespie, M. W. Rodger, Refreshing refreshable braille displays, IEEE transactions on haptics 8 (2015) 287–297.
[28] T. L. Varao Sousa, J. S. Carriere, D. Smilek, The way we encounter reading material influences how frequently we mind wander, Frontiers in psychology 4 (2013) 892.
[29] T. Götzelmann, L. Branz, C. Heidenreich, M. Otto, A personal computer-based approach for 3d printing accessible to blind people, in: Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments, 2017, pp. 1–4.
[30] B. Berman, 3-d printing: The new industrial revolution, Business horizons 55 (2012) 155–162.