

Media view shows any media content associated with the current text view, with alternative text support.

Navigation hint indicates to the reader what type of content is currently being displayed in the text and media views.

To Reading Accessibility and Beyond!

Our project highlights the high potential of immersive technologies to facilitate accessible media experiences far beyond simple magnification or image enhancement of classical assistive technologies. Their potential is supported by clinical proofs of concept, suggesting that smart augmentation principles might be beneficially implemented in VR [2], with the flexibility and full control over visual variables in a 3D environment, personalised to patients' needs. However, to leverage the VR/AR to design novel e-health systems, we need rigorous scientific, medical, and ergonomic evaluation of these systems.

This calls for translational research and cross-disciplinary approaches that implicate the vision science community to study fundamental and clinical vision science. An immediate challenge to confront here is the high technical barrier to create VR experiments. Existing packages like PsychoPy have successfully tackled this difficulty to generate stimuli on 2D monitors through a script programming interface. With the same ambition as PsychoPy, our team is currently working on the development of the Perception Toolbox for Virtual Reality (PTVR) package [L3], which proposes a scripting interface to design stimuli in virtual reality and is developed under an Open Science framework. This work is carried out in collaboration with E. Castet (Aix Marseille Univ, CNRS, LPC) in the context of ANR DEVISE [L4]. With the democratisation of immersive technologies and research-grounded solutions, our vision is that immersive technologies can soon be deployed on a larger scale to meet various patients' needs.

Links:

- [L1] <https://team.inria.fr/biovision/>
- [L2] <https://team.inria.fr/biovision/cardnews3d/>
- [L3] <https://ptvr.inria.fr/>
- [L4] <https://anr.fr/Projet-ANR-20-CE19-0018>

References:

- [1] H.-Y. Wu, A. Calabrèse, P. Kornprobst: "Towards Accessible News Reading Design in Virtual Reality for Low Vision", *Multimedia Tools and Applications*, Springer, 2021.
- [2] A. Calabrèse, et al.: "A Vision Enhancement System to Improve Face Recognition with Central Vision Loss", *Optometry and Vision Science*, 95(9), 2018.
- [3] M. Raphanel, et al.: "Current Practice in Low Vision Rehabilitation of Age-related Macular Degeneration and Usefulness of Virtual Reality as a Rehabilitation Tool," *J. Aging Sci.* 2018.

Please contact:

Hui-Yin Wu
Centre Inria d'Université Côte d'Azur, France
hui-yin.wu@inria.fr

Speech-to-Text Technology for Hard-of-Hearing People

by Manuela Hürlimann (Centre for Artificial Intelligence, Zurich University of Applied Sciences), Jolanda Galbier (Pro Audito Schweiz) and Mark Cieliebak (Centre for Artificial Intelligence, Zurich University of Applied Sciences)

Hard-of-hearing people face challenges in daily interactions that involve spoken language, such as meetings or doctor's visits. Automatic speech recognition technology can support them by providing a written transcript of the conversation. Pro Audito Schweiz, the Swiss federation of hard-of-hearing people, and the Centre for Artificial Intelligence (CAI) at the Zurich University of Applied Sciences (ZHAW) conducted a preliminary study into the use of Speech-to-Text (STT) for this target group. Our survey among the members of Pro Audito found that there is large interest in using automated solutions for better understanding in everyday situations. We now propose to take the next step and develop an application which uses ZHAW's high-quality STT models.



Figure 1: A group discussion – this is a situation in which the proposed application could support hard-of-hearing people (Photo: colourbox.de).

The average person holds more than 25 conversations per day, which can be very challenging for people with hearing loss, as their auditory perception of spoken language is limited. Pro Audito provides an interpreting service ("Schriftdolmetschen"), where a trained human interpreter accompanies the hard-of-hearing person and creates a written transcript of the interaction on the fly. While this is highly appreciated with 1,800 hours of speech transcribed each year, the financial compensation by the Swiss disability insurance is currently limited to professional and educational settings and the cost is capped [L1]. We received an Innovation Cheque from Innosuisse to run a preliminary study consisting of a needs analysis and market research. Our goal was to find out how STT could be used to create an offer for people with hearing loss that provides more flexibility and independence.

Needs analysis

The needs analysis was conducted via a detailed survey among the members of Pro Audito, which was answered by 166 respondents, of which 87% have moderate or severe hearing loss. We found that 28% already use technical support to facilitate understanding, which consists mostly of external microphones, headsets or rerouting sound to their hearing aid via Bluetooth (e.g., when watching TV). Some people already use

STT apps, where the most frequently named use cases are appointments at the doctor or optometrist, meetings (both online and on-site, see Figure 1) and conversations in crowded spaces with background noise (such as restaurants). 57% of our respondents can imagine using STT technology to facilitate their understanding – the most frequently named languages are Standard German, Swiss German, French and English. They were also asked what would be important features of an STT application: it should be as easy as possible to use and provide high-quality recognition (e.g., accuracy, robustness to noise, specialised vocabulary) with minimum latency. Many of our respondents would be willing to pay for a STT solution, either as a one-off purchase or on a monthly subscription basis. Most people would be willing to pay between 50 and 150 CHF one-off or 10 CHF per month.

Market Research

We reviewed existing STT solutions for people with hearing loss and found that currently no single solution ticks all the boxes – some have good recognition accuracy but a poor user interface, others are very easy to use but quickly become unstable when tested in real-life conditions. We are currently developing STT models for various languages at ZHAW. We believe that the best way forward is to develop a dedicated application for hard-of-hearing people and integrate our models for the following reasons:

- **Latency:** For real-time STT, latency needs to be minimised as much as possible. This means that ideally the model runs on-device, since using external cloud providers introduces an additional time-lag. Creating STT models which are small enough to run on a device such as a smartphone yet have high prediction accuracy is an important challenge.
- **Privacy:** Users will in some cases want to transcribe sensitive information, such as a conversation with a doctor. With a local model, privacy can be guaranteed.
- **Customisation:** The use cases from our survey offer significant challenges such as a large number of speakers, spontaneous speech, and background noise. If we use our own STT models, we have full control over their customisation.

Furthermore, it is important that this application can run on an inexpensive device to be accessible to as many users as possible; this is a further argument in favour of a smartphone app.

Future Activities

We propose to develop an application for hard-of-hearing people based on our STT models, which will use a high-precision microphone to record audio – either from the hearing aid itself, a partner microphone, or a wireless lapel microphone. The audio is then transmitted via Bluetooth to the user's smartphone. For minimum latency as well as maximum privacy and customisation, the transcription will be carried out on-device and will be displayed in an easy-to-use interface.

Pro Audito and ZHAW are now looking for partners interested in jointly developing and operating this application - if you are interested, please refer to the contact information below.

Link: [L1] <https://kwz.me/hjf>

Please contact:

Mark Cieliebak, ZHAW School of Engineering, Switzerland
ciel@zhaw.ch

Improving Accessibility for Deep Sky Observation

by Olivier Parisot, Pierrick Bruneau, Patrik Hitzelberger (Luxembourg Institute of Science and Technology), Gilles Krebs and Christophe Destruel (Vaonis)

Electronically Assisted Astronomy allows near-real-time generation of enhanced views of deep sky objects like nebulae and galaxies. This approach is ideal for people who have difficulty with direct observation through a telescope, especially those who have poor visual acuity or physical difficulties with positioning oneself correctly in front of the instrument.

Electronically Assisted Astronomy is widely applied today by astronomers to observe planets and deep faint sky objects like nebulae, galaxies, or globular clusters. By capturing images directly from a telescope coupled to a camera, this approach allows generation of enhanced views of observed targets that can be displayed in near real-time. While astrophotography aims at producing detailed and visually appealing images after numerous hours of post-processing of long exposure images [1], Electronically Assisted Astronomy aims at quickly getting images by stacking on-the-fly raw images in order to accumulate the (faint) signal (and then reduce the inherent noise). All this is made possible by the fact that recent CMOS/CCD cameras are extremely sensitive and have a very low read noise (i.e., amount of noise generated by electronics) [2], which makes it possible to obtain already satisfactory results with lightweight image processing.

By comparing with direct visual observation through an eyepiece and an instrument (refractor or reflector), this approach



Figure 1: Live session of Electronically Assisted Astronomy on the night of 14 May 2022 from a village in the northeast of France.