A speech translation system for medical dialogue in sign language — Questionnaire on user perspective of videos and the use of Avatar Technology

Bouillon, Pierrette; David, Bastien; Strasly, Irene; Spechbach, Hervé

Jekat, Susanne J., Steffen Puhl, Luisa Carrer, and Alexa Lintner, eds. 2021. *Proceedings of the 3rd Swiss Conference on Barrier-free Communication (BfC 2020), Winterthur (online), June 29–July 4, 2020.* Winterthur: ZHAW Zurich University of Applied Sciences. DOI: 10.21256/zhaw-3001.

Pierrette Bouillon, Bastien David, Irene Strasly & Hervé Spechbach

A speech translation system for medical dialogue in sign language — Questionnaire on user perspective of videos and the use of Avatar Technology

Research track

Abstract

This paper summarizes the findings of a questionnaire conducted with the Deaf community in francophone Europe. It aimed to gather feedback on specific features to be included in the sign language videos of BabelDr, a medical speech-to-sign translation platform. The results show that the whole sentence should be subtitled, the signer should be shown front-on only, and the background should be of a light colour. Explanatory images can be added to clarify the medical content. Lastly, human videos are preferred to avatars in this specific context, even if more than a third of the respondents think that an avatar could be useful in this context.

1 Deaf people's access to health care

WHO statistics indicate that one in every 1000 people in high-income countries is born deaf (Cantero 2016). These people are particularly vulnerable in terms of health. This vulnerability is due to several factors, including poor access to information and a lack of training for the medical staff. Another important factor in this context is the difficulty that they face in accessing care, both logistically (inability to make an appointment by telephone) and in communication (difficulty in understanding the caregivers, inadequate means of communication). These difficulties lead deaf patients to experience stressful and frustrating situations, which are frequently reported in the literature (Chastonay 2018).

The current legal framework in Switzerland sets conditions that make it easier for people with disabilities to participate in social life fully and equally. In particular, we here refer to the Federal law on the elimination of discrimination of persons with disabilities (DDA) that was enacted on December 13, 2002, as well as the UN Convention on the Rights of Persons with Disabilities (UNCRPD), which came into effect in 2014. Moreover, the Geneva canton has officially recognized sign language in its Constitution in June 2013.

In order to be in line with recent legal requirements, the BabelDr project aims to improve accessibility in hospitals. We developed a new type of fixed-phrase translator for triaging patients belonging to minority communities, including refugees, migrants and deaf people (Spechbach et al. 2019). Similar to other fixed-phrase translators (such as Medibabble or UniversalDoctor), the system relies on a predefined list of human-translated sentences in order to insure translation reliability, but instead of searching for sentences in this list with keywords, doctors can ask their questions orally, which improves the ergonomics (Spechbach et al. 2019).

This paper summarizes the findings of a questionnaire conducted to form an initial view of how Deaf people living in European francophone countries perceive the sign language videos elaborated for the BabelDr application, and the use of avatars in such videos (Chiriac et al. 2015). In particular, we wanted to know which factors affect their video quality preferences in order to decide how to present the videos in the BabelDr application.

In the next sections, first we briefly describe the BabelDr tool (Section 2) and how the sign language videos were realized (Section 3). We then focus on the questionnaire and the main findings of our research.

2 The BabelDr platform

BabelDr is an online speech-enabled fixed-phrase translator, specifically designed for medical dialogue (Spechbach et al. 2019). The system was developed based on four different criteria.

[Cost-efficiency] Machine Translation (MT) systems are not yet accurate enough to translate medical dialogues, especially into and from poorly supported languages (Patil and Davis 2014; Bouillon et al. 2017). For this reason, BabelDr is based on a set of pre-translated core sentences (including medical questions and instructions), organized by medical domains (type of pain). Although an ideal tool would have the flexibility of machine translation systems, different studies show that today fixed-phrase translation systems can be a good alternative to machine translation for this type of safety-critical domain (Turner et al. 2019).

[User-friendliness] To improve the flexibility and user-friendliness of this type of fixed-phrase translation tool, BabelDr uses speech recognition and neural methods to link the recognized sentence to the closest core sentence (Mutal et al. 2019). The core sentence is then translated for the patients by looking up in the translation memory. Although this type of fixed-phrase translator can be perceived as a constraint by doctors, the findings of the studies we made demonstrate that speech improves both satisfaction and usability (Boujon et al. 2018; Spechbach et al. 2019).

[Reliability] The reliability of the system is ensured by the back translation to the core sentence, which provides the doctor with the exact meaning of the spoken target sentence. The doctor must approve it before translating the sentence, meaning that they always know what will be translated.

[Flexibility] Content must be easily expandable. The translation memory of BabelDr contains around 13,000 core sentences and is constantly extended based on the system usage data. These core sentences are productively generated from a set of rules (patterns with variables) and mapped to more than one billion of source variations, which are used to train the speech recognizer and the neural matching algorithm. Human translation of core sentences can be done in different formats: written, spoken (for dialects), or video-based (for example, for sign languages). It is also possible to define different registers; for example, BabelDr translates specifically for female or male patients. The translation is done online in different ways, depending on the format. For written languages, the translators directly translate the patterns with a translation memory (Gerlach et al. 2018); If the translation is spoken or in the form of sign language videos, they translate the complete sentences.

At the time of writing, the BabelDr platform is accessible online at the address http://babeldr.unige.ch (see Figure 1) and is in use at Geneva University Hospitals (HUG) for translation between French and six migrant languages, with a high satisfaction from doctors and patients (Janakiram et al. 2020). French Sign Language (LSF) translation is currently under development, both with human and avatar videos (JASigning) in order to compare quality, satisfaction and development time, as described in more detail in the next section.

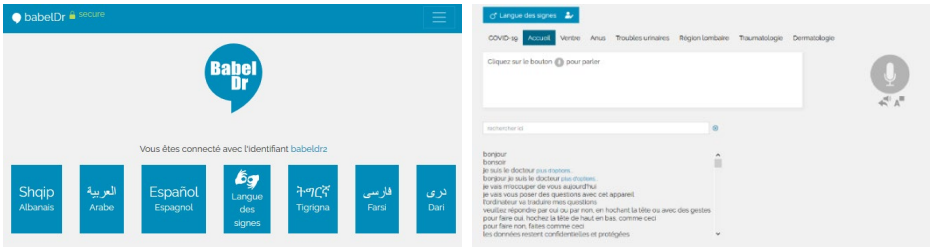

Figure 1. a) BabelDr Platform Home page, b) Doctor interface
© TIM/FTI, Geneva, 2020

3 Sign language videos

The LSF translation with human and avatars videos is currently being produced in two steps. A human translation is first carried out by a group of specialists (Deaf people and doctors) who work in a professional setting at the University of Geneva (Strasly et al. 2018). These human translations are then used as reference to produce the code for the JASigning avatar, using the grammar formalism described in Rayner et al. (2016). Figure 2 gives an example of human and avatar videos for the core sentence *Je suis l'infirmier* ("I am the nurse").

At present, 2,453 medical sentences have been human translated. The code used to generate sentences with JASigning avatar includes a glossary with 570 HamNoSys items used for describing manual signs (Smith 2013) and a grammar for linking glosses together and adding non-manual features.

During the human translation process, three main issues were raised by the translators: 1) translating medical terminology that did not have an exact equivalent in sign language; 2) translating using a video format (2D format) that does not allow the signer to make certain parts of the body easily visible, like the back; 3) translating of medicines (nouns): using the manual alphabet was not deemed appropriate by our translators since it would be very tiring for the person who would watch the videos. To face these issues, different solutions were adopted: 1) the use of paraphrases to explain the meaning of specific terminology, if that term had only one meaning that could be explained rapidly and in a clear way; 2) adding subtitles to the video and 3) adding explanatory images to give a clearer understanding of the sentence.

In the next sections, we describe the questionnaire aimed at gathering feedback on some of these different quality aspects (visual aspects, subtitles, and dimensionality), both for the human and the avatar videos.

4 Questionnaire methodology

4.1 Aims

The questionnaire aimed at gathering feedback on how Deaf people living in francophone European countries perceived specific elements of sign language videos produced as part of the BabelDr project, including the possible use of avatars.

4.2 Content

The questionnaire was conducted in French using the platform LimeSurvey. It was translated into French Sign Language (LSF), Swiss-French Sign Language (LSF-CH) and French Belgian Sign Language (LSFB) by professional translators, who were asked

to sign an informed consent form before being filmed.

It included 32 questions (Appendix) and covered six main themes: (i) background of the videos, (ii) additional images added to clarify the content, (iii) subtitles, (iv) screen format and size, (v) advertisements and logos displayed on the screen, and (vi) perception of the use of three-dimensional avatars. Each section consisted of three to six questions, not all of which were compulsory to answer. Our respondents first watched a short video summarizing the theme of the questions, then answered a series of Yes/No or multiple-choice questions (Haug et al. 2015).

The questionnaire was sent to potential participants from mid-October 2019 and closed at the end of November of the same year. In total, 61 associations in French-speaking European countries were contacted by email or via social networks (15 French-speaking Belgian associations, 39 French associations and 7 French-speaking Swiss associations). Private Facebook groups related to deafness or medical accessibility were also involved, sharing the link to the questionnaire with their members as part of a snowball sampling method.

5 Results

5.1 Respondents

A total of 111 people opened the questionnaire and started answering it: 47 LSF, 41 LSF-SR and 23 LSFB. Amongst them, thirty-three [N=33/111] gave consent for their data to be used. Of this group, 25 participants (13 LSF, 9 LSF-SR and 3 LSFB) completed the questionnaire.

34 comments were made by 13 different participants. Five main themes were addressed in the comments: “Avatars as an alternative means of communication” (7 items), “Comments on visual elements” (9), “Comments on Expressiveness” (4), “Adding written information or images” (15) and “Understanding signs” (7). We will report some of these comments in the original language (written French, sometimes with mistakes) alongside an explanation of their content in English.

Figure 2. “BE NURSE” signs: a) Human videos, b) Avatar videos.
© TIM/FTI, Geneva, 2020

5.2 Background of the video

54% of participants [N=18/33] considered the background of the video to be important for comprehension (Question 2.a [single-choice]). One of the participants said that when he answered he thought of those who have an usher syndrome: he suggested we avoid a white background or too many details on the background of the screen ("En ce qui concerne les fond écran, j'ai essayé de penser au usher donc à éviter le fond écran blanc ou trop de chose derrière"). Light colors (light grey) and plain backgrounds were clearly preferred (see also Zhao et al. 2000). As an example, one of the comments clearly stated that sometimes backgrounds hinder the understanding of what the avatar is signing because there is not enough contrast between the avatar and the background colours ("Parfois les arrière-plans empêchent une meilleure compréhension des propos de l'avatar en raison du contraste").

5.3 Additional images

54% of participants [N=18/33] considered explanatory images to be important for medical term comprehension, for example for locating an organ in the body (Question 3.a [single-choice]). To indicate a particular organ or the location of body pain, a vector graphic showing the whole body and a particular organ (42%, N=11/26) is preferred to anatomical drawings (31%, N=8/26) or a real picture (27%, N=7/26) (Question 3.b [single-choice]). Warm colours (orange, 57%, N=8/14, and red, 21%, N=3/14) are preferred for highlighting an organ in a vector image of a human body (Question 3.d [single-choice]). Explanatory images should be next to the signer and should take up no more than half of the screen (70%, N=14/20) so that signed information can be more clearly understood (see Figure 3.a).

5.4 Subtitles

86% of respondents (N=24/28) requested that subtitles be integrated into the videos to assist understanding of the signs (Question 4.a [single-choice]). Almost unanimously, respondents felt that the sentences should be fully subtitled (92%, N=22/24) and should be placed below the character (87%, N=21/24). One of our participants commented that adding subtitles below the signer could be useful for those who do not sign or who use Cued Speech ((…) "*ajouter des sous titre au dessous cela pourrait être utile pour les sourds qui ne signent pas ou qui utilise la LPC*").

Concerning the subtitle font colour, a white font with transparent background was preferred by the participants to other possibilities, for example black font with transparent background (33%, N=8/24 (Question 4.d [single-choice]) (see Figure 3.b).

For terms that are difficult to sign, respondents' opinions are divided between adding subtitles [N=19/24] and adding images [N=19/24] (Question 4.e [multiple-choice]). One of the participants stated the following:

> I was unsure whether I wanted to see more manual alphabet being used..: I would say yes, for those who find it difficult to read on the lips, but I think it is not so useful if there is already an image and a subtitle…knowing that images are stronger than anything --> DEAF PEOPLE = VISUAL PEOPLE ("j'étais partagée si oui ou non pour épeler en plus.. : je dirais oui pour ceux qui ont du mal à lire mais je pense que c'est pas très utile si il y a image + sous-titre .. sachant que l'image est plus fort que tout ---> SOURDS = VISUEL").

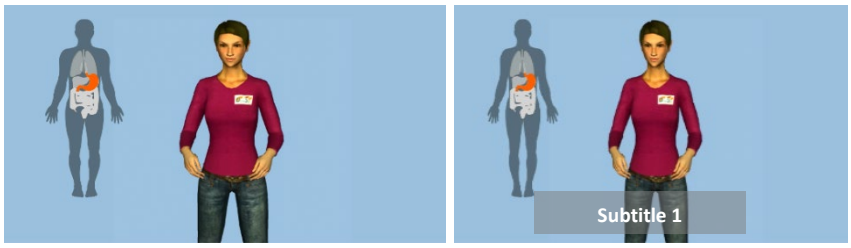

Figure 3. Selected images by participants: a) Explanatory picture, b) Incrusted subtitle

5.5 Screen format and size

Concerning the orientation of the screen, 87% of the participants [N=18/25] prefer a landscape view. The character should be shown from the waist up (80%, N=20/25) and when preparing the camera, a small space should be left above the head (84%, N=21/25) (Questions 5.d & 5.e [single-choice]). Images of the whole character are not appreciated and considered useless (see Figure 4.a).

5.6 Perception of three-dimensional avatars

A number of studies focussing on the perception of virtual avatars have shown that subjects experienced some difficulty in understanding the avatar (Huenerfauth et al. 2008; Kipp et al. 2011). Our results demonstrate that videos with real interpreters are more popular (64%, N=16/28) than the JASigning avatar, even if we were surprised to find that more than a third of the respondents think that an avatar could be useful in this context (36%, N=9/28). In particular, one comment that was added to this section states that avatars make it possible to avoid discrimination based on origin, age and gender ("(…) l'avatar est intéressant, car on peut choisir enfant, homme, femme, blanc, noir, etc. selon l'éthique auquel certains peuvent s'identifier sans aucune discrimination."). For the negative aspects, subjects had difficulty distinguishing between instructions and questions since the avatar was unable to use non-manual features to sign the difference ("Pour toute interrogation, on hausse les épaules quand il s'agit des questions", "je trouve qu'il manque des expressions faciale pour montrer que c'est une question"). Three other comments mention the robotic aspect of the avatar and the fact that they hope technology will improve ("Les avatars vous pouvez faire mieux, car ces avatars j'espère que c'est en test car je trouve comme les robots").

We also gathered feedback on how to the avatar should be displayed: 64% of the respondents [N=18/28] prefer the signer to be shown front-on only (Question 7.c [single-choice]). They generally do not want to see the avatar turned on the side, nor to have multiple perspectives of the same avatar on the same screen, as shown in Figure 4.b.

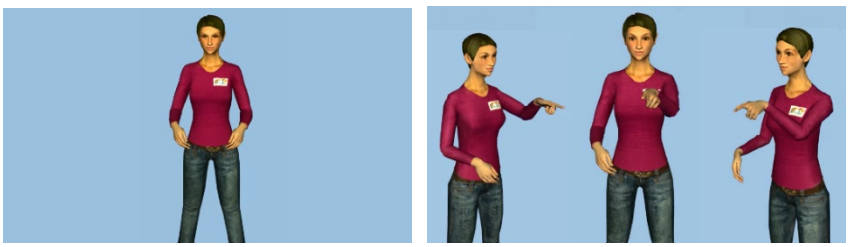

Figure 4. Non-selected images by participants: a) Full-length character, b) Multiple perspectives

6 Conclusion

In conclusion, the findings allowed us to gather the preferences of sign language users concerning the BabelDr tool. The results show that the whole sentence should be subtitled using a white font, the signer should be presented front-on only, and the background should be of a light colour. The signer should be filmed from the waist up above the head. Explanatory images can be added to clarify the medical content of the sentences. Human videos are preferred to avatars, but we were surprised to find that more than a third of the respondents think that a virtual human could be useful in this context (36%, N=9/28).

This experiment allowed us to present the project to the target audience and get their feedback on sign language videos and the use of an avatar. The next step will be to compare the comprehensibility of human and avatar videos in real emergency settings on a diagnosis task.

Acknowledgements

This research was made possible thanks to the support of Swissuniversities and the HUG Private Foundation. Irene Strasly was supported by the Swiss National Research Foundation.

References

Bouillon, Pierrette, Johanna Gerlach, Hervé Spechbach, Nikolaos Tsourakis, and Sonia Halimi. 2017. “BabelDr vs Google Translate: a user study at Geneva University Hospitals (HUG).” In *Conference Booklet of the 20th Annual Conference of the European Association for Machine Translation (EAMT)*, *Prague, Czech Republic, 2017*, 747–52. http://archive-ouverte.unige.ch/unige:94511.

Boujon, Valérie, Pierrette Bouillon, Hervé Spechbach, Johanna Gerlach, and Irene Strasly. 2018. “Can Speech-Enabled Phraselators Improve Healthcare Accessibility? A Case Study Comparing BabelDr with MediBabble for Anamnesis in Emergency Settings.” In *Proceedings of the 1st Swiss Conference on Barrier-free Communication*, *Winterthur, Switzerland, 2018,* 50–65. Winterthur: ZHAW digitalcollection. https://archive-ouverte.unige.ch/unige:105852.

Cantero, Odile. 2016. “Accès aux soins et communication : vers une passerelle entre la communauté sourde et les soignants de Suisse romande.” PhD diss., University of Lausanne, Switzerland.

Chastonay, Oriane, Joëlle Blanchard, Odile Cantero, Magali Carmes, Pascaline Michel, Pascal Singy, Patrick Bodenmann, and Benoit Mongourdin. 2018. “Les Sourds: une population vulnérable méconnue des professionnels de la santé.” *Forum Médical Suisse – Swiss Medical Forum* 18 (38): 769–74. https://doi.org/10.4414/fms.2018.03361.

Chiriac, Ionut Adrian, Lacramioara Stoicu-Tivadar, and Elena Podoleanu. 2015. “Comparing Video and Avatar Technology for a Health Education Application for Deaf People*.” Studies in Health Technology and Informatics (HTI)*: 516–20.

Elliott, Ralph, John Glauert, Richard Kennaway, Ian Marshall, and Eva Sáfár. 2008. “Linguistic Modelling and Language-Processing Technologies for Avatar-Based Sign Language Presentation.” *Universal Access in the Information Society* 6 (4): 375–91. https://doi.org/10.1007/s10209-007-0102-z.

Gerlach, Johanna, Hervé Spechbach, and Pierrette Bouillon. 2018. “Creating an Online Translation Platform to Build Target Language Resources for a Medical Phraselator.” In *Proceedings of the 40th edition of Translating and the Computer Conference (TC40)*, 60–65. London, Geneva: AsLing, The International Association for Advancement in Language Technology. https://archive-ouverte.unige.ch/unige:111776.

Haug, Tobias, Rosalind Herman, and Bencie Woll. 2015. “Constructing an Online Test Framework, Using the Example of a Sign Language Receptive Skills Test.” *Deafness & Education International* 17 (1): 3–7. https://doi.org/10.1179/1557069X14Y.0000000035.

Huenerfauth, Matt, Liming Zhao, Erdan Gu, and Jan Allbeck. 2008. “Evaluation of American Sign Language Generation by Native ASL Signers.” *ACM Transactions on Accessible Computing* 1 (1): 1–27. https://doi.org/10.1145/1361203.1361206.

Janakiram, Antony Akash, Johanna Gerlach, Alyssa Vuadens-Lehmann, Pierrette Bouillon, and Hervé Spechbach. 2020. “User Satisfaction with a Speech-Enabled Translator in Emergency Settings.” In *Digital Personalized Health and Medicine*, edited by Louise B. Pape-Haugaard, Christian Lovis, Inge Cort Madsen, Patrick Weber, Per Hostrup Nielsen, and Philip Scott, 1421–22. Amsterdam, Berlin, Washington D.C.: European Federation for Medical Informatics (EFMI), IOS Press.

Kipp, Michael, Alexis Heloir, and Quan Nguyen. 2011. “Sign Language avatars: animation and comprehensibility.” In *Intelligent Virtual Agents*, edited by Hannes Högni Vilhjálmsson, Stefan Kopp, Stacy Marsella, et Kristinn R. Thórisson, 6895:113–26. Berlin, Heidelberg: Springer.

Mutal, Jonathan, Pierrette Bouillon, Johanna Gerlach, Paula Estralla, and Hervé Spechbach. 2019. “Monolingual backtranslation in a medical speech translation system for diagnostic interviews - a NMT approach.” In *Proceedings of Machine Translation Summit XVII* (Volume 2: Translator, Project and User Tracks), *Dublin, Ireland, 2019,* edited by European Association for Machine Translation: 169–203.

Rayner, Manny, Pierrette Bouillon, Sarah Ebling, Johanna Gerlach, Irene Strasly, and Nikos Tsourakis. 2016. “An Open Web Platform for Rule-Based Speech-to-Sign Translation.” In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Volume 2: Short Papers), *Berlin, Germany, 2016.* 162–68. https://doi.org/10.18653/v1/P16-2027.

Smith, Robert, ed. 2013. HamNoSys 4.0: User guide. Blanchardstown: Institute of Technology. https://robertsmithresearch.files.wordpress.com/2012/10/hamnosys-user-guide-rs-draft-v3-0.pdf.

Spechbach, Hervé, Johanna Gerlach, Sanae Mazouri Karker, Nikolaos Tsourakis, Christophe Combescure, and Pierrette Bouillon. 2019. “A Speech-Enabled Fixed-Phrase Translator for Emergency Settings: Crossover Study.” *JMIR Medical Informatics* 7 (2): e13167. https://doi.org/10.2196/13167.

Strasly, Irene, Tanya Sebaï, Evelyne Rigot, Valentin Marti, Jesus Manuel Gonzalez, Johanna Gerlach, Hervé Spechbach, and Pierrette Bouillon. 2018. “Le projet BabelDr : rendre les informations médicales accessibles en Langue des Signes de Suisse Romande (LSF-SR).” In *Proceedings of the 2nd Swiss Conference on Barrier-free Communication*, *Geneva, Switzerland, 2018,* 92–96. https://archive-ouverte.unige.ch/unige:114719

Patil, Sumant, and Patrick Davies. 2014. “Use of Google Translate in Medical Communication: Evaluation of Accuracy.” *BMJ* 349 (December): g7392–g7392. https://doi.org/10.1136/bmj.g7392.

Turner, Anne, Yong Choi, Kristin Dew, Ming-Tse Tsai, Alyssa Bosold, Shuyang Wu, Donahue Smith, et Hendrika Meischke. 2019. “Evaluating the Usefulness of Translation Technologies for Emergency Response Communication: A Scenario-Based Study.” *JMIR Public Health and Surveillance* 5 (1): e11171. https://doi.org/10.2196/11171.

Zhao, Liwei, Karin Kipper, William Schuler, Christian Vogler, Norman Badler, and Martha Palmer. 2000. “A Machine Translation System from English to American Sign Language.” In *Envisioning Machine Translation in the Information Future*, edited by John S. White, 54–67. Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-39965-8_6.

Authors

Pierrette Bouillon has been Professor at the FTI ("Faculty of Translation and Interpreting"), University of Geneva since 2007. She is currently Director of the Department of Translation Technology and Dean of the FTI. She has numerous publications in computational linguistics and natural language processing, particularly within lexical semantics (Generative lexicon theory), speech-to-speech machine translation for limited domains, post-editing and more recently accessibility. She is now head of the new SNF Propicto project, "PRojection du langage Oral vers des unités PICTOgraphiques – PROPICTO".

Email: pierrette.bouillon@unige.ch
Website: https://www.unige.ch/fti/en/faculte/departements/dtim/membrestim/bouillon/

Bastien David is a Research Assistant at the Translation Technology Department of the Faculty of Translation and Interpreting at the University of Geneva since November 2018. He has some publications in speech-to-sign machine translation for medical and transport domains. As part of the BabelDr project, he is working on the creation of virtual characters that will automatically translate from French into the Sign Language of French-speaking Switzerland (LSF-CH).

Email: bastien.david@unige.ch
Website: https://www.unige.ch/fti/en/faculte/departements/dtim/membrestim/david/

Irene Strasly holds a Master in specialized translation delivered by the Faculty of Translation and Interpreting in Geneva. She joined the FTI staff in 2013. Previously, she worked as teaching assistant at the Italian Unit, and then she started to collaborate with the Department of Translation Technology (TIM). Together with Prof. Pierrette Bouillon, she currently coordinates various sign language projects related to sign language and accessibility. Since February 2017, she is working on a thesis on sign language accessibility in the French-speaking area of Switzerland.

Email: irene.strasly@unige.ch
Website: https://www.unige.ch/fti/en/faculte/departements/dtim/membrestim/strasly/

Hervé Spechbach is MD at the University Hospitals of Geneva (HUG) since 2006 and head of the emergency service since 2015. He has many publications in the field of informatics for healthcare. He is representing the University Hospitals of Geneva in the Swiss Centre for Barrier-free Communication project and co-leads the BabelDr project, an innovative and reliable medical phraselator, which received an innovation award from the University Hospitals of Geneva in 2018.

Email: herve.Spechbach@hcuge.ch