Article

User Experience Questionnaire in sign language for native   
users of Slovenian Sign Language

Klemen Pečnik 1\*, Žana Juvan 1, Gregor Dolinar 2, and Matevž Pogačnik 1,\*

|  |
| --- |
| **Citation:** Lastname, F.; Lastname, F.; Lastname, F. Title. *Sensors* **2022**, *22*, x. https://doi.org/10.3390/xxxxx  Academic Editor: Firstname Lastname  Received: date  Accepted: date  Published: date  **Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.    **Copyright:** © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/). |

1 Laboratory for Multimedia, Faculty of Electrical Engineering;

2 Laboratory of Applied Mathematics and Statistics, Faculty of Electrical Engineering;

**\*** Correspondence: klemen.pecnik@fe.uni-lj.si; matevz.pogacnik@fe.uni-lj.si; Tel.: +386 31 758 885 K.P.; +386 1 4768 101 M.P.

**Abstract:** Accessibility of interactive services and applications is an important part of an inclusive society, even more so in the light of the ongoing digitalisation process. While legal and regulatory frameworks are being implemented and the services are adopting accessibility standards, we still lack user experience evaluation approaches which would be adapted to deaf participants. Therefore, the standard User Experience Questionnaire (UEQ) was translated to Slovenian Sign Language (SSL) and its appropriateness and reliability evaluated in a research study. For this purpose, two interactive TV applications for deaf viewers were developed and evaluated using 1) a standard written UEQ in Slovenian language (UEQ TXT) and 2) a UEQ translated to Slovenian Sign Language (UEQ SSL). The evaluation was done by 27 deaf participants and 6 Slovenian Sign Language interpreters. A thorough statistical analysis of the results was performed on the level of individual UEQ statements as well as on the level of UEQ scales. The results are very promising and show that there were no statistically significant differences in evaluation results when gathered with the UEQ SSL compared to results gathered with UEQ TXT. The participants also expressed a strong preference for UEQ SSL as it allows deaf to more independently participate in user experience evaluation studies.

**Keywords:** UEQ, User Experience, User Testing, Accessibility, Deaf and Hard of Hearing, Deaf Community, Sign Language, Slovenian Sign Language, interactive multimedia, HbbTV

1. Introduction

The development of assistive technologies has significantly contributed to quality-of-life improvements for the deaf and hard of hearing, especially in the fields of communication and education. There are many organisations contributing to such improvements, with one of most important ones being W3C Web Accessibility Initiative (WAI), standardizing the Web Content Accessibility Guidelines (WCAG) [1]. Many countries are striving towards improvement of the accessibility situation, mostly through legal and regulatory frameworks, while there is still a lot of room for improvement in terms of funding the development of accessibility solutions. The Republic of Slovenia is an EU member state and is therefore following EU directives, enacting accessibility laws and gradually progressing towards accessible public services. The World Health Organisation (WHO) has projected that by 2050 nearly 2.5 billion people will have some degree of hearing loss [2], while at the moment there are 75 million deaf people in the world. In Slovenia there are around 1500 deaf people with 1000 of them using the Slovenian Sign Language as their primary language [3]. While a laudable improvement on accessibility regulations, technological solutions, and service development has been made in the past 10 years, a lack of user-friendly approaches adapted to deaf participants for user testing and user experience evaluations is still noticeable. A growing awareness that the evaluations and testing of new technologies and services are of paramount importance has resulted in many new evaluation methods and approaches, however these are adapted only to the general population without disabilities. In A Case Study about Usability, User Experience and Accessibility Problems of Deaf Users with Assistive Technologies [4] researchers were evaluating the usability and user experience of assistive technologies and one of the research findings was the conclusion that users are not adopting assistive technologies due to bad user experience. We believe that bad user experience with assistive technologies as well as any accessible solutions could be significantly improved if the evaluation procedures themselves are made accessible and are adapted for the deaf population.

Sign languages are non-verbal languages, based on concepts and depending a lot on facial expressions and body movements in addition to hand movements. They have specific grammatical rules and syntax, and thus often differ from spoken languages. The structure of the sentences in sign languages is simplified for faster communication, the word order is different, and declensions and conjugations do not exist. This means that spoken languages cannot be literally translated into sign languages and must be contextually interpreted. In general, sign languages have smaller vocabularies in comparison to spoken languages, i.e. the Slovenian Sign Language has 20.000 recognized signs [5] while the Slovenian spoken language has more than 110.000 words [6]. Nevertheless, this does not imply that sign languages are not fully developed. Even more, they are considered by linguists as full-fledged and often even more expressive than spoken languages, with dialects and different social genres of language appearing even within an individual sign language [7]. A common misconception is that sign language is international or that all sign languages are the same worldwide. Like spoken languages, sign languages can also be classified into language families. However, countries with the same spoken language do not necessarily have the same sign language e.g., American and British Sign Language developed independently of each other and are therefore not mutually intelligible. For communication on an international level, there exists International Sign (IS) which is often used at international conferences and events. It is a simplified form of a sign language with a limited set of words and does not have the status of the language, as it does not have a precisely defined vocabulary [8]. In Slovenia the Slovenian Sign Language has been recognized as a native language and as such added to the Slovenian constitution on June 4th, 2021 [9]. Sign language interpreters bridge the communication gap between native users of sign language and the hearing community in various situations. In Slovenia, similarly to many other countries, a sign language interpreter is a profession gained with special examination in front of the National Profession Qualification Committee for sign language interpreters, certifying highly skilled knowledge and qualification. There were 64 registered SSL interpreters in the registry at the time of conducting the study and many among them are Children Of Deaf Adult(s) (CODA). All other users of sign language without qualification are typically called signers, they can communicate conversationally with people who are deaf and hard of hearing but do not have the necessary official qualifications [10,11].

There is a plethora of quantitative and qualitative research methods that help the developers with evaluations of different solutions’ user experience and its aspects, such as usability, task difficulty, effort, mental demand, learnability, attractiveness, novelty, trust, VR sickness, etc. The present study is mainly focused on the quantitative methods as the goal is to get an objective insight into the user experience of proposed solutions with an emphasis on multimedia content distribution and informational services. One of the oldest and widely accepted methods is the System Usability Scale (SUS), which is quick and easy to use, and is often used as a benchmark method [12]. It consists of a 10-item questionnaire with a 5-point scale answers, resulting in a single score ranging from 0 to 100 indicating the usability and learnability of the evaluated solution. A similar method to SUS, but somewhat more targeted and with more sub scores, is the Post Study System Usability Questionnaire (PSSUQ) [13]. It is a 19-item questionnaire with a 7-point scale and provides scores for system usefulness, information quality, interface quality, and an overall score. Similarly to SUS, it can be done using either a paper or a digital version. Another well-known method is NASA Task Load Index (NasaTLX) which rates perceived workload to assess a task, system, or team’s effectiveness [14]. It consists of 21 item questionnaire, 15 pairing questions and 6 scales, giving three different results. It is more complex than SUS and quite demanding to set up. Measurements of software quality from the users’ perspective is often performed using the Software Usability Metric (SUMI) methodology [15]. It has been used for over 25 years and consists of 50-item questionnaire with a 3-point scale, providing results for efficiency, affect, helpfulness, control, learnability, and a global SUMI score. It produces reliable results with small sample sizes and allows for their comparison with the global SUMI repository of over 2000 evaluated software solutions. However, a 50-item questionnaire can be quite tedious for the participants to solve and requires a lot of participants’ time and effort. On the other hand, The Psychosocial Impact of Assistive Devices Scales (PIADS) was designed to assess the effects of an assistive device. The self-report questionnaire consists of 26 self-report items with 7-point Likert scale, giving results on functional independence, well-being, and quality of life [16] PIADS is translated into 15 written languages and originating in English. While PIADS is focusing on assistive devices, the present study is more focused on user interfaces and services. Therefore, the User Experience Questionnaire (UEQ) [17,18] was selected as a target evaluation method since it is a quick and reliable method and includes both classical usability measurements as well as different user experience aspects. It consists of a 26-item questionnaire with a 7-point Likert scale and returns scores on 6 scales: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty, thus covering the pragmatic and hedonistic aspects of evaluated solutions. It also contains many useful tools to help calculate the results’ statistical reliability, identify inconsistencies, calculate scores using benchmark values, and finally produce reports.

It is apparent that measuring user experience (UX) is a general practice with the goal of creating a satisfying experience for users of products, services, and designs. There are many different UX methods, however they are mainly based on the existing and standardized models, making them hard to apply in certain use cases. This is especially true in the case of trying to evaluate UX with the deaf and hard of hearing, due to the general assumption that all deaf can read and write. About 0.1% of the population becomes deaf before learning any language, consequently a sign language is their native language and as such the main form of communication. Therefore, deaf individuals are not familiar with the phonetics of spoken languages, they have difficulties with understanding, and they often find written words hard to read as they usually do not understand the structure of the sentences and other linguistic peculiarities of spoken languages [19,20]. In this aspect, standardized questionnaires for measuring UX are not a viable option, which limits the used methods to discussions with sign language interpreters present [21] or simple approaches like using a smile-o-meter [22]. Since UX evaluations have to include the actual target users when developing user interfaces, services etc. for the deaf and hard of hearing, alternative approaches of measuring UX have to be used [23] and whenever one of the standard methods is applied, there is a risk of getting the wrong data due to misunderstandings. Although there are many studies and research work available in the field of user experience, assistive technologies, and accessibility, not a lot of research can be found on accessible user experience evaluation tools. In order to provide equal rights and opportunities the main goal of the present study was to adapt the UEQ, a standardized method for measuring user experience, for all native users of Slovenian Sign Language regardless of their degree of hearing loss or identification with the deaf community.

The research goal of the present study was to evaluate the appropriateness and usefulness of a UEQ questionnaire translated into the Slovenian Sign Language (SSL) [24]. If proven suitable, the translated questionnaire would enable more independent evaluations of interactive services by the native users of SSL, with minimized need for a sign language interpreter’s involvement. This has not only positive practical consequences but also a high symbolic significance. Namely, it would allow for more independent involvement of deaf users, which is one of the important principles of the “United Nations Convention on the Rights of the Persons with Disabilities” [25].

The pilot testing of the UEQ questionnaire in Slovenian Sign Language was performed using two interactive solutions: 1) A Hybrid Broadcast-Broadband Television (HbbTV) application with an adjustable sign language interpreter layout and 2) A virtual 3D sign language interpreter for a limited vocabulary, which has been recognized as a solution for enhancing accessibility for broadcast services [26] and also implemented in the form of an HbbTV application. HbbTV as a platform has also been recognized as the technology that may be used to enhance accessibility services [27]. The experiment details are described in the Materials and methods section.

The results are very encouraging and present a basis for future work in this field.

1. Materials and Methods

In this section the approach to the SSL translation of the standardized UEQ questionnaire is explained, followed by a description of the apparatus, designed for the evaluation experiment, and the description of the experiment procedure. Finally, the demographic structure of the experiment’s participants is described.

* 1. Augmentation of the standardized UEQ with sign language

The questionnaire translations, experiments, measurements, and final analysis are based on the standardized textual UEQ in Slovenian language. The questionnaire consists of 26 questions, which are not actual questions, but rather sets of two opposing statements. The users decide which of the opposing statements suits their experience best, rated on a 7-point Likert scale. All gathered responses are then used in the official UEQ Data Analysis Tool, where they are transformed and the main results are presented on 6 scales: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty.

An extensive study involving several SSL interpreters has been conducted, with the goal of finding the best matches and most appropriate sign language translations for all 26 UEQ statement pairs. The translation process has been conducted in accordance with the guidelines of the UEQ Handbook [28] and Collaborative and Iterative Translation approach as proposed by Douglas, S. P., & Craig, C. S. (2007) [29]. All written UEQ statement pairs were first evaluated by the team of two SSL interpreters and four members of the Laboratory for Multimedia, Faculty of Electrical Engineering (LMMFE), using multiple standardized UEQ translations (Slovenian, English and German) to find the statement's best equivalent in SSL, considering its scale, concept, and functional meaning. For 18 statement pairs it was possible to find a direct translation from written Slovenian to Slovenian Sign Language. However, for the remaining 8 pairs a direct translation was not possible, because those expressions do not exist as signs in the SSL vocabulary. An example of such statement is the statement "dull", which was translated contextually to the signed expression "without ideas". Another challenge when translating the statements was the fact that a singular UEQ statement should only appear once in the whole questionnaire, since similar written expressions often translate in the same sign (gesture). Such examples are expressions “conventional” and “usual” or “leading edge” and “innovative”. In accordance with the selected translation methods a list of all statements in SSL was created and all signs were recorded with the SSL interpreter, who is also a CODA, in a professional broadcasting studio in front of a green screen, as shown in Figure 1 and Figure 2.

All recorded statements were first evaluated with the mentioned interpreter and some of the signs were recorded in multiple versions. After basic postproduction the videos of the recorded signs were presented to another SSL interpreter, who is a mid-schoolteacher in a school for deaf children, a DODA (Deaf Of Deaf Adult(s)), and additionally a certified International Sign interpreter. He translated all signs back to Slovenian words, which were then revised and discussed by the whole team. This process has been repeated two times in order to make all statements as straightforward as possible. There was a dilemma about the translation of two statements in SSL since there exist multiple gestures with the same meaning. The difference between them is that some are officially confirmed and included in the SSL dictionary, while others are used colloquially and usually more often than the official ones. Upon the interpreters’ suggestion the officially recognized gestures were used, as the evaluation tool would become official once validated. Documentation of development, changes, team comments, and user comments in the preliminary testing has been established and will be maintained in further testing and validation process.

|  |  |
| --- | --- |
|  | A person standing in front of a group of people sitting at desks  Description automatically generated with low confidence |
| **Figure 1.** Second iteration of the translation and recording  process | **Figure 2.** Final iteration of the translation and recording process within the pilot study |

The main goal of the present study was to provide an accessible UEQ translated to Slovenian Sign Language (UEQ SSL) as a standalone solution in the form of videos in a purpose-built web page, allowing for very simple interactions. All recorded statements were edited in post-production, the green background was removed and the videos containing corresponding statement pair signs were placed in their respective positions as presented in Figure 3. Two versions of such statement pair videos were exported, in the first one both videos of the recorded statements were playing simultaneously, and in the second one they were playing sequentially.

* 1. Apparatus

The whole testing procedure was originally designed as a Hybrid Broadcast Broadband Television (HbbTV) application with the backend system working over digital video broadcasting – terrestrial (DVB-T) network. However, due to the lack of availability of HbbTV 2.0 compliant TV sets [30], a web-based application running in Chrome web browser with the RedOrbit HbbTV emulator extension [31] was implemented, allowing for a simple and intuitive user interaction using a regular TV remote control. As the usage of new technologies could be questionable, the television was used as it is one of the most used technologies across deaf and hard of hearing. The goal was to simplify the user interface and user interaction as much as possible, from this perspective the TV remote control is one of the most recognised and commonly accepted interfaces. Considering all comments and suggestions from the preliminary evaluations, user comments from the interviews, and recommendations for sign language interpreter positioning and sizing [32,33] the whole research framework (experiments and evaluations) was designed as an application with a predefined structure and workflow and with the assumption that most of the potential users would be deaf or hard of hearing. All available navigation options were always shown on top of the application in the form of a standardized HbbTV legend, explaining how the standardized buttons on all smart TV remote controls can be used.

The UEQ evaluation part of the experiment was designed as a simple animated questionnaire with intuitive navigation and the same design approach for all evaluations. The pre-recorded videos of SSL interpreters signing the statement pairs were shown in full screen for each respective UEQ question, with a simple 7-point scale positioned in the middle of the screen and a progress bar placed at the bottom to indicate the user’s progress in the evaluation. To select a value on the 7-point scale for each statement pair the participants used the arrow buttons on the TV remote and clicked OK to confirm their choice. The option to return to any previous question at any point of time was also added, allowing for easy mistake correction during the testing. Users were also able to easily switch between simultaneous and sequential signing, with the sequential signing being the default option. This decision was based on the results and comments from the preliminary testing where it was pointed out that while some users might be ok with simultaneous signing it is usually too demanding for most people in terms of concentration. During the preliminary study the UEQ comprised of both signed and written expressions. However, due to the fact that one of the main goals was to assess the understandability and appropriateness of the UEQ SSL variant, separate UEQ SSL and standard written UEQ in Slovenian (UEQ TXT) evaluation versions were created.

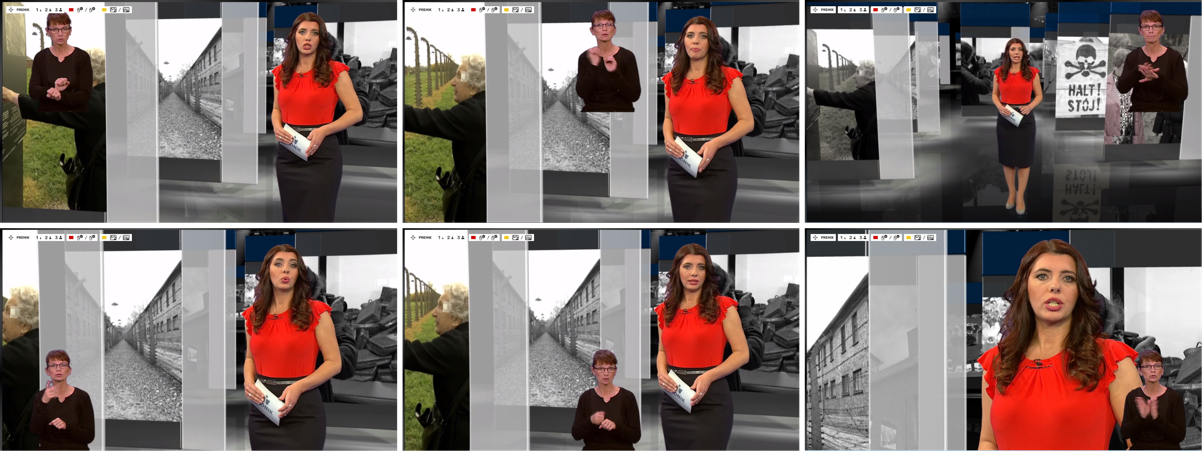
Graphical user interface

Description automatically generated

**Figure 3.** UEQ SSL (left) and UEQ TXT (right) for the statements ‘fast – slow’

An example for the statements fast – slow is presented in Figure 3, with the SSL version on the left side and the standard version on the right side. There are many studies and approaches for pilot testing and validation process of translated questionnaires, while for the present study three main options were identified as applicable. In the first option the participants would first respond to the UEQ TXT and then to the UEQ SSL, however during the preliminary study a drawback of this approach was identified as the participants were biased by the originating questionnaire (UEQ TXT) causing constant back-comparison of target language, the SSL version, to the original questionnaire. The second option with randomly alternating both versions of questionnaires was at first recognized as the best one. However, the UEQ handbook as well as the guidelines for the validation of the translated questionnaires used as measuring instrument suggest that non-alternating methods should be used in such cases. After consultation with the Association of Deaf Teachers in Slovenia, the option of alternating the UEQ versions was considered for the validation of the UEQ SSL for future work. Meanwhile during methodology development, the method of the preliminary psychometric testing of the pre-ﬁnal version of the translated instrument with a bilingual sample has been recognized as the most appropriate approach. In accordance with “The participants respond to the items of the pre-final version of questionnaire in target language (P-FTL) without seeing the original instrument in the source language (SL). After completion of the P-FTL, the participants are given the original instrument in the SL and are asked to answer the items” [34], UEQ SSL has been used first, followed by the standard UEQ TXT statements in the pilot study.

As aforementioned, two solutions were developed for the pilot testing of the UEQ SSL. During the first solution, the adjustable interpreter, users were able to adjust the position of the SSL interpreter on the screen by pressing the left-up-right-down arrow keys. This allowed the users to move the interpreter out of the way in case it was covering an important part of the screen, without sacrificing the size of the original video. As presented in Figure 4, there were 6 predefined positions available. The content for the solution was acquired from the national broadcaster Radio-television Slovenia (RTV SLO) as it was broadcasted live over the DVB-T networks with the difference that the main video and the SSL interpreter video were in this case separated, thus allowing to provide different variants. While there are recommendations on sign interpreters positioning, size and background [35,36], the initial size, positioning and type of overlay were intentionally the same as they are on the broadcasted national TV programme.



**Figure 4.** Sign interpreter positions

By pressing the numbers 1, 2 and 3, the users were able to adjust the size of the overlayed interpreter, with 1 being the smallest and 3 being the largest option, as presented in Figure 5.



**Figure 5.** Sign interpreter sizes (from left to right: small, medium, big)

The users also had the option to switch off or turn on the sign interpreter video by pressing the red button on the TV remote control as a demonstration of how the presented solution could be switched on and off on a smart TV while watching a regular TV channel. Last but not least, the layout mode could be changed from "overlay" to "side-by-side" by pressing the yellow button on the TV remote control, as presented in Figure 6.

A collage of people

Description automatically generated with low confidence

**Figure 6.** Video and sign interpreter layout (on the left an overlay example and on the right a side-by-side example)

The second interactive application that was used for testing the UEQ SSL was a virtual 3D sign language interpreter for a limited vocabulary and for selected domains of use, like emergency notifications, traffic announcements, schedule changes, delays, or as in the case of the pilot study the domain of weather forecast. The virtual interpreter was developed in the scope of another project and is in the present study used only as an example of an interactive service, intended for user experience evaluation. This application allowed for the same functionalities and options as the first one, allowing the participants to adjust the virtual interpreter's position, size, and layout. Additionally, the option to toggle subtitles was added by pressing the blue button on the TV remote control. The participants were able to evaluate the virtual interpreter's understandability, attractiveness, novelty, etc. using the UEQ method. Figure **7** shows examples of two different layouts of a virtual interpreter for the weather forecast domain. Based on initial preliminary studies it was expected that the UEQ results for the adjustable interpreter could be too good so the second experiment, a virtual interpreter, with lower acceptance level and consequently with lower UEQ results expected was added to the pilot testing. There is a plethora of research projects on designing and using virtual sign language interpreters as well as studies on acceptance and attitude of targeted groups. Based on recent studies on acceptance and attitude towards virtual sign language interpreters, showing that virtual interpreters are still not fully accepted by deaf and hard of hearing community, a virtual interpreter has been selected to intentionally acquire worse results of user experience [37]. For the UEQ SSL it is important to be accurate in good and bad user experience evaluation results, and in this perspective second solution with virtual interpreter was added. Both developed solutions were used only as a service to be evaluated with the UEQ SSL within the pilot study and as such are not a subject of the study itself.

Graphical user interface

Description automatically generated

**Figure 7.** Examples of two different layouts for a virtual sign language interpreter

Furthermore, the introduction, brief experiment explanation, and UEQ instructions have also been prepared in sign language as well as in written text and were included as videos, creating a seamless user experience throughout the evaluation. An example of the instructions is shown in Figure 8.

Graphical user interface

Description automatically generated

**Figure 8.** UEQ instructions in Slovenian Sign Language with subtitles

* 1. Experiment procedure

User testing was conducted in the DemoLab of the Laboratory for Multimedia, at the Faculty of Electrical Engineering, University of Ljubljana in cooperation with the Association of Deaf Teachers in Slovenia.

The room provided a comfortable environment for the participants, which were sat on a sofa in front of a 55” OLED TV (LG OLED55C6V). User interaction with interactive application and the UEQ has been limited to the use of a TV remote control from the beginning to the end of the experiment, as shown in . The whole experiment was designed to provide a home living experience to relax the participants, make them feel comfortable and without inducing the feeling that they are a subject of the testing.

A picture containing text, indoor, person

Description automatically generated

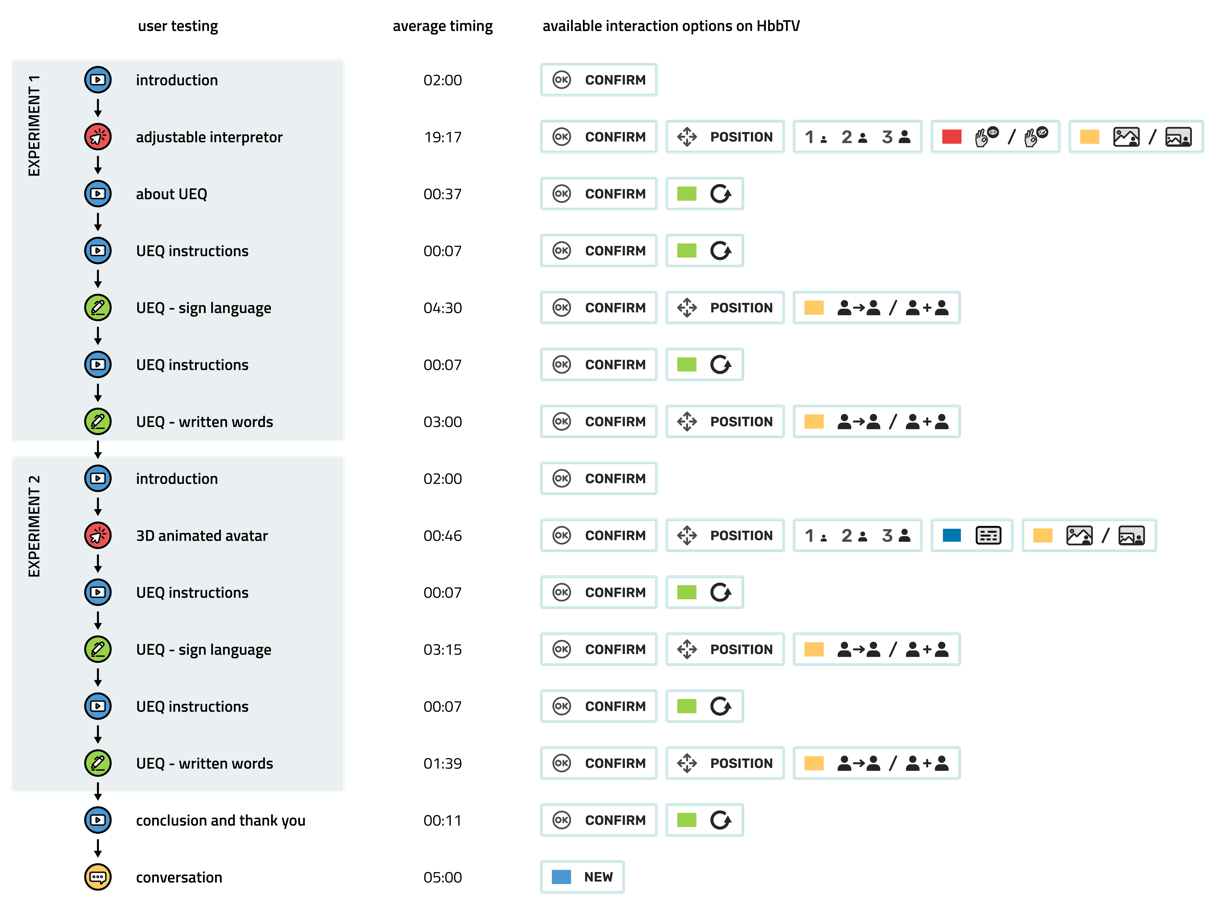
**Figure 9.**

At the beginning of each evaluation the participant would fill in a simple questionnaire with the purpose of gathering demographic data, information about their technical experience and other relevant information, and sign a consent form for the purpose of the evaluation. Each test started with a simple presentation of the experiment idea and an explanation of all interaction possibilities. At the beginning of each of the two experiments (adjustable and virtual interpreter), the participant was given a couple of tasks and was asked to:

1. change the position of the interpreter,
2. change the size of the interpreter,
3. change the layout,
4. turn the interpreter off and then back on (since the interpreter was already present at the beginning of the experiment),
5. adjust the interpreter size, position, and layout according to their personal preference.

After a participant finished testing the interactive application of adjustable interpreter, a video with an explanation of the UEQ structure and answering procedure was presented. Then the participant first answered the UEQ SSL, followed by the UEQ TXT. This procedure was repeated for both experiments. After each step in the experiments the application waited for the participant's confirmation before continuation, and the users always had the option to repeat the last step. At the end of the second experiment the participants filled a post-test questionnaire, followed by a discussion which was conducted in a manner of a relaxed conversation. All comments and suggestions about both experiments were gathered, along with the opinion about the involvement of deaf users in user testing. The whole step-by-step process with all the steps, average completion timings, and all available user interactions is presented in Figure 10.

Despite the pre-recorded instructions and explanations, a sign language interpreter was present during the testing especially with deaf participants to clarify any additional questions and to translate all given comments and suggestions to and from the SSL during the discussion. Additionally, two user experience experts were present at each test session. One expert was responsible for leading the evaluation procedure, answering potential questions, and providing additional explanations, while the other was taking notes about observed difficulties and challenges that a participant may have had and noted all relevant comments during the discussion phase. The collection of notes in conjunction with user action logs have allowed for a comprehensive analysis of proposed solutions. Additionally, they have given a lot of useful information on the issues and dilemmas related to the chosen sign language expressions, their understandability, and appropriateness.



**Figure 10.**

At the beginning of each user test a random user ID was generated and displayed in the upper right corner of the screen. The ID was used to associate each participant’s UEQ answers, questionnaires, and notes, without the need to identify them by their names or any other personal data, thus allowing anonymisation. The data for all four UEQ questionnaires was automatically stored in an SQL (Structured Query Language) database immediately after each keypress, recording both the answer and the time of completion, allowing for further data analysis. All responses were exported to a .csv format and transferred into the UEQ Data Analysis Tool [17] as well as to an additional Excel spreadsheet for further analysis.

* 1. Demographic data

Most participants’ responses were gathered during the Deaf History International 2022 conference, organised by the Association of Deaf Teachers in Slovenia, Association "Theater, Audiovisual Arts and Culture of the Deaf - DLAN", and Deaf History International in Ljubljana and Zagreb [38].

In the pilot testing only native SSL users with at least basic reading skills were suitable to participate. The study is an ongoing research work in cooperation with the deaf community including associations of the deaf and interpreters. Therefore, the number of participants is changed every month. In the pilot study 36 participants took part in the experiment, with 47 % being female and 53 % male participants, as shown in Figure 11. Age structure was gathered in age ranges with 6 % of participants aged 15-25 years, 17 % 26-35 years, 19 % 36-45 years, 36 % 46-55 years, and 22 % aged 56 or more, as shown in Figure 12.

|  |  |
| --- | --- |
|  |  |
| **Figure 11.** Gender distribution of the participants | **Figure 12.** Age structure of the participants |

Out of all participants 53 % were deaf from birth, 31 % lost hearing later in life, 17 % had no hearing impairment, and no participants were hard of hearing, as shown in Figure 13. Furthermore, all who lost hearing later explained that they lost hearing in early childhood and are native SSL users. Out of the 17 % of hearing participants 11 % were certified SSL interpreters and CODA, 3 % certified SSL interpreter and 3 % teachers in a school for deaf children with more than 10 years of SSL teaching experience but not certified as an interpreter. Because UEQ SSL is addressing all users of SSL, despite deafness type, cause or level of hearing loss, additional data regarding details about types of deafness and levels of hearing loss was not obtained.

|  |  |
| --- | --- |
|  |  |
| **Figure 13.** Type of deafness | **Figure 14.** Participants' education levels |

Since the educational level could have an influence on the understanding of the written language, information about the participants’ education was also collected. 6 % of all participants finished only primary school, 50 % secondary school, 14 % short-cycle higher vocational education (SCHVE), 8 % obtained a bachelor’s degree, 17 % a master’s degree, and 6 % a PhD, as shown in Figure 14. Finally, 8 % of participants' results were excluded from the analysis, since the results had too many data inconsistencies and therefore had to be eliminated as recommended in the UEQ instructions. All three participants were identified as having lost hearing later, one from the age group 36-45, one from 46-55, and one from 56 or more.



1. Results

In all standard user experience evaluations with the UEQ questionnaire the pragmatic and hedonistic UEQ values represent the measure of success or failure. In this case, however, the UEQ values themselves are of secondary importance, as the appropriateness of the UEQ SSL is evaluated. Therefore, the results are presented in the form of a comparison between the UEQ TXT questionnaire values and the UEQ SSL questionnaire values.

For this purpose, the following Theorems were defined:

**Theorem 1.** The responses obtained with the UEQ TXT and the UEQ SSL will be similar without significant statistical differences.

**Theorem 2.** The results of the evaluation using the UEQ TXT and the UEQ SSL will give comparable results in all UEQ scales (Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty).

**Theorem 3.** Deaf users will prefer the user experience evaluation procedure when using the UEQ SSL or a combination of both as compared to the evaluation when using the UEQ TXT only.

* 1. Statistical analysis of the gathered responses

First, all gathered data from the experiments was evaluated and the statistical analysis was performed. The raw data was then further evaluated using the official UEQ Data Analysis Tools. In some cases minor anomalies were spotted and they are presented and explained with additional information in the Discussion.

The subtraction difference between the values of the answers from UEQ SSL and UEQ TXT was calculated, with the maximum possible difference between them being 6 or -6, and the target value being 0. The analysis which was focused on the answers for each UEQ question shows that the best match between standard UEQ TXT and UEQ SSL was obtained for UEQ question 11 with 30 exact matches out of 33 possible and the worst match was obtained for question 8 with 18 exact matches. When focused on each participant, the analysis shows that the best result was 26 exact matches out of 26 possible, meaning that 3 participants answered both UEQs equally, and the worst result was 11 exact matches from 2 participants. The distribution of all differences between UEQ SSL and UEQ TXT for adjustable interpreter is presented in Figure 15.

**Figure 15.** Differences between UEQ TXT and UEQ SSL distribution N=858

However, the UEQ results are always presented using UEQ scales and calculated accordingly, thus the distribution of differences within each UEQ scale has been analysed and, as presented in Figure 16, the distributions of differences for each UEQ scale are very similar. Distribution of differences in percentages is gathered in Table 1.

**Figure 16.** Distribution of differences between UEQ TXT and UEQ SSL per scales N=858

**Table 1**. Distribution of differences between UEQ TXT and UEQ SSL per scales N=858

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **difference** | Attractiveness | Perspicuity | Efficiency | Dependability | Stimulation | Novelty |
| -6 | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % |
| -5 | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % |
| -4 | 0 % | 0 % | 0 % | 0 % | 0 % | 2 % |
| -3 | 0 % | 0 % | 0 % | 0 % | 2 % | 1 % |
| -2 | 1 % | 1 % | 0 % | 3 % | 2 % | 0 % |
| -1 | 8 % | 14 % | 11% | 13 % | 11 % | 10 % |
| 0 | 73 % | 69 % | 70 % | 68 % | 73 % | 68 % |
| 1 | 18 % | 11 % | 15 % | 12 % | 9 % | 14 % |
| 2 | 1 % | 4 % | 4 % | 2 % | 3 % | 4 % |
| 3 | 0 % | 1 % | 0 % | 2 % | 0 % | 2 % |
| 4 | 0 % | 0 % | 0 % | 0 % | 1 % | 0 % |
| 5 | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % |
| 6 | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % |

Using the same principle, the differences distribution of the virtual interpreter was analysed resulting in almost the same distribution with an even higher peak for differences of 0.

While the differences seem to be quite consistent in distribution manner, paired t-test has been used to check if answers in text and sign language are the same. Question pairs were defined as vectors from UEQ SSL and UEQ TXT, as shown in (1).

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

The paired t-test has two alternative hypotheses, i.e., the null hypothesis assumes that the mean of difference of paired answers in text and sign language is equal to zero, while the alternative hypotheses implies that there is a significant difference between the two populations. In equation for t-value (2) represents the observed mean of responses for one question in textual form, represents the observed mean of responses for the same question in sign form (corresponding pairs - e.g. ), represents sample standard deviation of the differences of textual and sign pair responses, and represents the sample size.

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

All t-values for individual questions were compared with the corresponding T-value for a = 0.05 with 32 degrees of freedom, which corresponds to the value of 2.042 [39], and all P-values with T.DIST.2T function were calculated. We rejected the null hypothesis *H0* for questions 6 and 9, due to the P-value being 0.032 for question 6, and 0,005 for question 9, and concluded from alternative hypotheses *Ha* that the responses to questions 6 and 9 are significantly different, while for all the other questions that is not the case.

To detect irregularities, user errors, or misunderstood signs, an in-depth analysis of each separate question is useful, however, in the present study the statistical analysis of results of UEQ scales is more important.

As mentioned, the six UEQ scales are obtained by calculating the mean of responses to the corresponding questions. Attractiveness is determined by 6 questions, while all other scales (Perspicuity, Efficiency, Dependability, Stimulation, and Novelty) are determined by 4 questions. For this reason, the vectors were defined as:

|  |  |  |
| --- | --- | --- |
|  |  | () |

with

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

where each of the vectors and for corresponds to one of the UEQ scales with cardinality of, i.e.,, equal to 6 for Attractiveness, and with cardinality of equal to 4 for all other UEQ scales. An appropriate statistical test is needed to compare the responses for both scales. In this manner both samples were tested one against another by using the Multivariate Paired Hotelling's T-Square test [40] with the null hypothesis

|  |  |  |
| --- | --- | --- |
|  |  | () |

and the alternative hypothesis

|  |  |  |
| --- | --- | --- |
|  |  | () |

where are mean vectors for vectors and respectively.

To compare the two corresponding samples, vectors Yi are defined as:

|  |  |  |
| --- | --- | --- |
|  |  | () |

and

|  |  |  |
| --- | --- | --- |
|  | . | () |

Subsequently, the null hypothesis that the population mean vector is equal to 0 can be tested, which would mean that all of its elements are equal to 0. This can be tested against the alternative that the vector is not equal to 0, which would mean that at least one element of the vector is not equal to 0:

|  |  |  |
| --- | --- | --- |
|  |  | () |

The sample mean vector was defined as:

|  |  |  |
| --- | --- | --- |
|  |  | () |

and the matrix with the sample variance-covariance of the vectors as:

|  |  |  |
| --- | --- | --- |
|  |  | () |

Finally, a paired Hotelling’s T-Square test is given with the expression:

|  |  |  |
| --- | --- | --- |
|  |  | () |

and the corresponding F-statistic is defined with the expression

|  |  |  |
| --- | --- | --- |
|  |  | () |

where and represent the degrees of freedom and can be rejected at level  if the F-value exceeds the value from F-value with and degrees of freedom, evaluated at level . Results of the Paired Hotelling’s T-Square test for all UEQ scales for the adjustable interpreter are presented in Table **2** and for the virtual interpreter in Table 3.

**Table 2.** Results of the Paired Hotelling’s T-Square test for all UEQ scales for the adjustable interpreter

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **UEQ Scales** | n | p | df1 | df2 | T2 | F-value | p-value |
| Attractiveness | 33 | 6 | 6 | 27 | 15.423 | 2.169 | 0.101 |
| Perspicuity | 33 | 4 | 4 | 29 | 2.794 | 0.633 | 0.643 |
| Efficiency | 33 | 4 | 4 | 29 | 9.331 | 2.114 | 0.108 |
| Dependability | 33 | 4 | 4 | 29 | 1.510 | 0.342 | 0.847 |
| Stimulation | 33 | 4 | 4 | 29 | 10.410 | 2.359 | 0.080 |
| Novelty | 33 | 4 | 4 | 29 | 5.310 | 1.203 | 0.333 |

**Table 3.** Results of the Paired Hotelling’s T-Square test for all UEQ scales for the virtual interpreter

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **UEQ Scales** | n | p | df1 | df2 | T2 | F-value | p-value |
| Attractiveness | 32 | 6 | 6 | 26 | 4.827 | 0.675 | 0.616 |
| Perspicuity | 32 | 4 | 4 | 28 | 8.209 | 1.854 | 0.149 |
| Efficiency | 32 | 4 | 4 | 28 | 6.284 | 1.419 | 0.256 |
| Dependability | 32 | 4 | 4 | 28 | 8.246 | 1.862 | 0.147 |
| Stimulation | 32 | 4 | 4 | 28 | 10.301 | 2.326 | 0.083 |
| Novelty | 32 | 4 | 4 | 28 | 1.872 | 0.423 | 0.791 |

**Proof of Theorem 1.** The results of the Paired Hotelling’s T-Square test show that we cannot reject the null hypothesis for each of the UEQ scales since all p-values are larger than 0.05. This indicates that the responses in UEQ TXT and UEQ SSL are not significantly different. □

* 1. Analysis and comparison of the UEQ results for both experiments

Firstly, the calculated results from the UEQ Data Analysis Tool were obtained for the UEQ SSL along with calculated results from the UEQ TXT. In order to determine the differences of mean values, variances, standard deviations, and confidence intervals between both UEQ versions, differences of mean values for each UEQ scale were compared, with the results presented in Table 4 with the UEQ SSL results in the 2nd column, standard UEQ TXT results in the 3rd column and the differences in the 4th column.

UEQ results of both approaches can be identified as significantly close as the absolute differences in values for individual UEQ scales range between 0.008 (Stimulation) and 0.116 (Attractiveness), with the highest relative difference around 5% (Attractiveness). This seems to be a good result, especially considering that the UEQ scale values are in practice interpreted as excellent, good, above average, below average, or bad, which further reduces the significance of such minor differences in result values.

**Table 4.** Comparison of UEQ Mean values for the adjustable interpreter

|  |  |  |  |
| --- | --- | --- | --- |
| **UEQ Scales** | **Mean SSL** | **Mean TXT** | **Δ Mean** |
| Attractiveness | 2.646 | 2.530 | 0.116 |
| Perspicuity | 2.515 | 2.462 | 0.053 |
| Efficiency | 2.402 | 2.288 | 0.114 |
| Dependability | 2.311 | 2.288 | 0.023 |
| Stimulation | 2.545 | 2.553 | -0.008 |
| Novelty | 2.500 | 2.417 | 0.083 |

Similarly, to the procedure of the adjustable interpreter, mean values were calculated and compared for the virtual interpreter which are presented in Table 5.

**Table 5:** UEQ Mean values for the virtual interpreter

|  |  |  |  |
| --- | --- | --- | --- |
| **UEQ Scales** | **Mean SSL** | **Mean TXT** | **Δ Mean** |
| Attractiveness | 2.052 | 2.036 | 0.016 |
| Perspicuity | 2.047 | 1.903 | 0.117 |
| Efficiency | 2.016 | 1.938 | 0.078 |
| Dependability | 1.789 | 1.609 | 0.180 |
| Stimulation | 2.000 | 2.039 | -0.039 |
| Novelty | 2.234 | 2.273 | -0.039 |

Again, the results of both UEQversions were comparable, with the absolute difference in values for individual UEQ scales being between 0.039 (Stimulation and Novelty) and 0.180 (Dependability). This also seems to be a good result, especially considering that UEQ scale results are in practice interpreted as excellent, good, above average, below average, or bad, as presented in Figure 17.

**Figure 17.** UEQ SSL results for virtual interpreter

The UEQ Data Analysis Tool also includes embedded tools for detecting data inconsistencies and for calculating scale consistencies. Data inconsistencies are reported when differences between the best and the worst answer values on the same scale exceed 3 points. If such a difference appears in three or more scales, the participant’s data is marked as critically suspicious and should be eliminated. For both experiments 3 cases of critically suspicious data were detected and consequentially eliminated from the analysis. Most inconsistencies were identified in the Dependability scale.

The scale consistency is checked against the whole scale by calculating an average of correlation of all answers in the UEQ scale with Cronbach’s Alpha-Coefficient and Gutman’s Lambda2 Coefficient. Cronbach’s Alpha-Coefficient is calculated as:

|  |  |  |
| --- | --- | --- |
|  |  | () |

where represents the sample size (6 for Attractiveness and 4 for other scales) and represents an average of all correlations between sample items. As it can be noticed from the equation, the sample size is fixed (6 or 4) and the actual correlation has an influence on the coefficient value. Guttman’s Lambda2 coefficient for each UEQ scale is calculated as:

|  |  |  |
| --- | --- | --- |
|  | , | () |

with cardinality of, i.e.,, equal to 6 for Attractiveness, and with cardinality ofequal to 4 for all the other UEQ scales. Variable represents the number of questions for the corresponding scale (6 for Attractiveness and 4 for other scales), represents the variance of the jth question, represents the variance of the observed scores on the composite measures for corresponding question, and represents the covariances between items *i* and *j*.

Using the rule of thumb, it is recommended that the Cronbach’s Alpha-Coefficient and the Gutman’s Lambda2 coefficient are higher than 0.7. In case of the adjustable interpreter these coefficient values are above the recommended threshold value for all the scales except in the case of Efficiency (= 0.66) and Dependability (= 0.41) for the UEQ SSL, and in the case of Dependability (a = 0.67) for the UEQ TXT. This result could either be a consequence of having a smaller sample size of 33 participants in the study or a consequence of corresponding statement understandability. Namely, 14 % of the participants had issues with understanding the statements for question 8, and more than 42% of participants did not exactly understand the user experience context of statements 17 and 19. The results of scale consistency for Cronbach’s Alpha and Gutman’s Lambda2 are presented in Table 6.

**Table 6.** UEQ scale consistency of the adjustable interpreter



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **UEQ Scales** | **SSL** | **TXT ** | **SSL** | **TXT ** |
| Attractiveness | 0.79 | 0.86 | 0.80 | 0.85 |
| Perspicuity | 0.75 | 0.83 | 0.76 | 0.80 |
| Efficiency | 0.66 | 0.82 | 0.59 | 0.78 |
| Dependability | 0.41 | 0.67 | 0.40 | 0.63 |
| Stimulation | 0.63 | 0.77 | 0.59 | 0.73 |
| Novelty | 0.74 | 0.68 | 0.69 | 0.65 |

In the experiment with the virtual interpreter the scale consistency results were all above the recommended threshold value of 0.7. The results are presented in .

**Table 7.** UEQ scale consistency of virtual interpreter

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **UEQ Scales** | **SSL a** | **TXT a** | **SSL l2** | **TXT l2** |
| Attractiveness | 0.96 | 0.98 | 0.95 | 0.98 |
| Perspicuity | 0.88 | 0.96 | 0.87 | 0.96 |
| Efficiency | 0.87 | 0.87 | 0.78 | 0.88 |
| Dependability | 0.87 | 0.87 | 0.84 | 0.84 |
| Stimulation | 0.96 | 0.96 | 0.95 | 0.95 |
| Novelty | 0.87 | 0.84 | 0.85 | 0.83 |

In addition, the variances and the standard deviations for both UEQ versions as well as the standard deviations of their differences were calculated for both experiments. For the adjustable interpreter the standard deviation of the answer value differences for each individual UEQ question varies between 0.30 and 1.05, while the standard deviations of answer value differences belonging to each UEQ scale are presented in Table 8. The 2nd column represents the standard deviation of UEQ SSL scales, the 3rd column a standard deviation of UEQ TXT scales and the 4th column a standard deviation of differences between UEQ TXT and UEQ SSL evaluation.

**Table 8.** Standard deviation of differences between UEQ TXT and UEQ SSL per scale for the adjustable interpreter

|  |  |  |  |
| --- | --- | --- | --- |
| **UEQ Scales** | **Std. dev. SSL** | **Std. dev. TXT** | **Std. dev. TXT-SSL** |
| Attractiveness | 0.452 | 0.452 | 0.000 |
| Perspicuity | 0.616 | 0.586 | -0.030 |
| Efficiency | 0.570 | 0.562 | -0.008 |
| Dependability | 0.508 | 0.616 | 0.107 |
| Stimulation | 0.454 | 0.478 | 0.024 |
| Novelty | 0.618 | 0.580 | -0.038 |

As the UEQ evaluation results are usually interpreted on the level of each scale, these results are more significant than each answer value individually. The standard deviation of the UEQ SSL scale results varies from 0.452 (Attractiveness) to 0.618 (Novelty) and of the UEQ TXT from 0.452 (Attractiveness) to 0.616 (Dependability). Meanwhile the absolute differences vary from 0.000 (Attractiveness) to 0.107 (Dependability). It can be noticed that the highest calculated values of standard deviation are for Novelty in UEQ SSL and for Dependability in UEQ TXT and differences between both versions.

In the same manner as for the first experiment the variance and all standard deviations were calculated for the virtual interpreter. The results are presented in Table 9.

**Table 9.** Standard deviation of differences between UEQ TXT and UEQ SSL per scale for the adjustable interpreter

|  |  |  |  |
| --- | --- | --- | --- |
| **UEQ Scales** | **Std. dev. SSL** | **Std. dev. TXT** | **Std. dev. TXT-SSL** |
| Attractiveness | 1.164 | 1.311 | 0.147 |
| Perspicuity | 1.061 | 1.140 | 0.079 |
| Efficiency | 0.959 | 0.893 | -0.065 |
| Dependability | 1.221 | 1.178 | -0.043 |
| Stimulation | 1.208 | 1.298 | 0.090 |
| Novelty | 0.888 | 0.896 | 0.008 |

The standard deviation of the UEQ SSL scale results vary from 0.888 (Novelty) to 1.221 (Dependability) and of the standard UEQ from 0.893 (Efficiency) to 1.311 (Attractiveness). Meanwhile the absolute differences vary from 0.008 (Novelty) to 0.147 (Attractiveness).

Finally, to address the Theorem2, a comparison of means and confidence intervals of the UEQ scales for the adjustable interpreter are presented in graphical form in Figure 18.

**Figure 18.** Comparison of UEQ results between UEQ TXT and UEQ SSL for the adjustable interpreter

The same comparison of the UEQ TXT and UEQ SSL for the virtual interpreter can be observed in Figure 19.

**Figure 19.** Comparison of UEQ results between UEQ TXT and UEQ SSL for the virtual interpreter

The values of means and their corresponding confidence intervals are in both cases coequal, except for the Dependability scale where, as already mentioned, participants had issues with understanding some of the statements.

**Proof of Theorem 2.** The comparison of the UEQ scale results between UEQ TXT and UEQ SSL have shown that they were significantly close. Due to the representation and interpretation of the UEQ results it can be concluded that the UEQ SSL would return comparable results as UEQ TXT. □

* 1. Evaluation of participants’ preference of the UEQ version

Considering the analysis of the demographic data and comments from the participants relating to the UEQ results it can be concluded that a UEQ SSL would mean a lot to the participants with lower education, especially for those who could not read. Only one of the participants did not find the UEQ SSL beneficial, while all other participants stated that they would prefer the UEQ SSL or a combination of both to the standard UEQ TXT. The participants also stated that they would prefer to be independent during such activities and consequently sign language evaluations are needed and welcome.

**Proof of Theorem 3.** During the interviews and discussions 95 % of participants preferred the use of the UEQ SSL (or a combination of UEQ SSL with added text) to the use of the UEQ TXT only. □

1. Discussion

The main goal of the presented research was to confirm the working hypothesis that the results of the evaluations using the UEQ TXT and the UEQ SSL will give comparable results in all scales (Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty). During the preliminary evaluations of two different interactive solutions intended for the deaf and hard of hearing population a sufficiently small UEQ scale result differences were obtained between the UEQ SSL and the UEQ TXT. Furthermore, the standard deviations of corresponding answers’ value differences on the level of the UEQ scales are also relatively small. Therefore, the results of the UEQ SSL and the UEQ TXT can be claimed to be comparable. However, considering the analysis and the origin of the differences, the way how the results are calculated, and the sample size obtained, it can be expected that with the sample size of obtained group above 50 it would be sufficient to get reliable results with all scale consistency coefficients above 0.7.

It should also be noted that very high UEQ values of both evaluated solutions can mostly be contributed to the lack of such solutions in real life and consequently the deaf are gratefully embracing any solution that would help them in everyday life. Other information obtained during the interviews with participants includes a high praise for both the developed applications which were evaluated as well as for the evaluation procedure using a sign language version of the UEQ questionnaire. Namely, all deaf participants have expressed the wish to participate in the development of such interactive solutions. Therefore, the need to use sign language-based evaluation tools seems even greater and more relevant, which is based on the proof of the Theorem3, stating that deaf users will prefer the user experience evaluation procedure when using UEQ SSL.

Other interview observations, given by seven deaf participants, include various suggestions for alternative sign language translations of individual UEQ statements. Despite conducting an extensive study with a study group for the UEQ translation to SSL, choosing only official signs during the preliminary study and pilot testing, there were some additional suggestions to change particular sign interpretations for the UEQ statements ‘inventive’, ‘conventional’, ‘motivating’, ‘organized’, and ‘cluttered’. On the other hand, participants understood statements ‘secure’ and ‘not secure’ or ‘unpredictable’ and ‘predictable’ but did not understand their meaning in the context of user experience. For statements ‘meets expectations’ and ‘does not meet expectations’ 36 % of participants stated that they understood the statements but claimed they did not have any expectations of the evaluated applications. Besides that, it was noticed that for the UEQ questions whose meaning was not clearly understood the participants tended to answer randomly which decreased the statistical significance of the data. With questions 6, 8, and 9 such cases were noticed, also resulting in very low corelation between other UEQ scale corresponding questions, and in scale consistency coefficients under 0.7.

Based on the results of the pilot study it would be recommended to use a combined UEQ questionnaire (statements in text and sign language together) as a UEQ evaluation tool for the native users of SSL after the final version is validated. The analysis of simultaneous vs. sequential playback of UEQ statement pairs has shown that only 5 % of the participants preferred the simultaneous option, while all other participants stated that they prefer the sequential playback of signed pair statements, noting that it is easier to focus on one video of a signed statement at the same time.

The results of the pilot study have been satisfactory and have given useful directions for future work and further validation of the UEQ SSL. First and foremost, we plan to extend the study with additional participants with an updated experiment to validate the translation and obtain statistically even more reliable results. Additionally, some of the UEQ SSL signs will be modified as suggested by some of the participants and evaluated by the working group for the UEQ translation. As the UEQ TXT is already translated in 36 written languages, it would be of a great importance to also translate it in other sign languages accordingly. Finally, both evaluated interactive applications show great promise for their usefulness and wider adoption by the deaf community and should be developed further with consideration of guidelines for positioning sign language interpreters [41,42]. Finally, the target audience for UEQ SSL are all deaf and hard of hearing whose native language is the Slovenian Sign Language, although during the pilot testing we were unable to obtain hard of hearing participants, but it does not mean that they should be excluded in any way.

* 1. Limitation of the study

To validate the UEQ SSL the sample size should be significantly increased and the translated UEQ tested on more solutions. Nevertheless, concerns about biased result have arisen due to the chosen method of preliminary psychometric testing of the pre-ﬁnal version of the translated instrument with a bilingual sample, which could be addressed with an additional method of alternating versions of UEQ TXT and UEQ SSL, which would be used during testing with a doubled sample size in order to detect and eliminate bias of each approach.

Last but not least, the present study was postponed as a result of covid restrictions and health concerns, which on the other hand resulted in early conclusion. Study had to be concluded due to time limitations of deadlines before reaching the target sample size of 50 participants.

1. Conclusions

The domain of the presented research is becoming increasingly relevant as the web and digital interactive solutions in general are becoming key elements of communication and everyday life. As an estimated 80 million people in the EU alone live with a disability, it is more necessary than ever to ensure that everyone has equal access to digital products and services. The EU Web Accessibility Directive (Directive (EU) 2016/2102) [43,44] is a step in this direction, which aims to consolidate accessibility standards, making web accessibility a legal requirement. Consequently, the present study addressing the adaptation of evaluation tools for deaf and hard of hearing seems to be a step in the right direction. Furthermore, with a few improvements of the evaluation method presented in this study such evaluations could be used for different services and use cases on a number of platforms. As such, accessible user experience evaluations could include a wider population and decrease the need of a sign interpreter presence with every participant during evaluations. Furthermore, just like the UEQ translations to spoken languages are gradually increasing, using the proposed method the questionnaire could be translated in many other sign languages, consequently increasing the inclusiveness of the deaf and hard of hearing.

**Author Contributions:** “Conceptualization, Klemen Pečnik, Matevž Pogačnik and Žana Juvan; methodology, Klemen Pečnik and Matevž Pogačnik; software, Klemen Pečnik and Žana Juvan; validation, Klemen Pečnik and Žana Juvan; statistical analysis, Gregor Dolinar and Klemen Pečnik; resources, Klemen Pečnik and Žana Juvan; data curation, Klemen Pečnik and Gregor Dolinar; writing—original draft preparation, Klemen Pečnik, Matevž Pogačnik and Žana Juvan; writing—review and editing, Klemen Pečnik, Matevž Pogačnik and Žana Juvan; visualization, Žana Juvan; supervision, Matevž Pogačnik; project administration, Klemen Pečnik; All authors have read and agreed to the published version of the manuscript.”

**Funding:** “This research received no external funding”

**Institutional Review Board Statement:** “The study was conducted in accordance with the Declaration of Helsinki, and in accordance with the Guidelines for ethical conduct in research involving people, defined by the ethics committee of the University of Ljubljana.”

**Informed Consent Statement:** “Informed consent was obtained from all subjects involved in the study.”

**Data Availability Statement:** Raw experiment data from the SQL database and demographic data (scanned, and only in Slovenian language) available upon request.

**Acknowledgments:** The authors would like to sincerely thank the Association of Deaf Teachers in Slovenia and Association of Deaf Slovenia 14 11 for their help within the study and in promoting this research and gathering of study participants. This research was in part supported by Slovenian Research Agency (ARRS), P2-0425: Decentralized solutions for the digitalization of industry and smart cities and communities.

**Conflicts of Interest:** “The authors declare no conflict of interest.”

References

1. Initiative (WAI), W.W.A. WCAG 2 Overview Available online: https://www.w3.org/WAI/standards-guidelines/wcag/ (accessed on 16 August 2022).

2. World Health Organization Deafness and Hearing Loss Available online: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss (accessed on 16 August 2022).

3. ZDGNS Gluhost Available online: http://zveza-gns.si/o-zvezi/o-gluhoti/ (accessed on 16 August 2022).

4. Godoi, T.; Silva Junior, D.; Valentim, N. A Case Study About Usability, User Experience and Accessibility Problems of Deaf Users with Assistive Technologies. In; 2020; pp. 73–91 ISBN 978-3-030-49107-9.

5. Varuh človekovih pravic RS Varuh Svetina v luči mednarodnega dneva o uresničevanju pravic gluhih in naglušnih Available online: https://www.varuh-rs.si/index.php?id=1965&L=tseexcwihoggpxrj&tx\_news\_pi1%5Bnews%5D=6252&tx\_news\_pi1%5Bcontroller%5D=News&tx\_news\_pi1%5Baction%5D=detail&cHash=143ccdd6c11be6b04b2d635baa64545b (accessed on 22 September 2022).

6. Ramovš, F. Slovar slovenskega knjižnega jezika 2021.

7. Schlenker, P. Visible Meaning: Sign Language and the Foundations of Semantics. *Theor. Linguist.* **2018**, *44*, 123–208, doi:10.1515/tl-2018-0012.

8. International Sign Available online: https://www.eud.eu/about-us/eud-position-paper/international-sign-guidelines/ (accessed on 25 July 2018).

9. RTV Slovenija Slovenski znakovni jezik je del jezikovne dediščine Slovenije Available online: https://www.rtvslo.si/dostopno/clanki/slovenski-znakovni-jezik-je-del-jezikovne-dediscine-slovenije/583831 (accessed on 21 September 2022).

10. ZZTSJZ Poklicna kvalifikacija tolmač Available online: https://www.tolmaci.si/poklicna-kvalifikacija-tolmac/ (accessed on 22 September 2022).

11. NECC What Is a Sign Language Interpreter? Available online: https://facstaff.necc.mass.edu/faculty-resources/deaf-and-hard-of-hearing-services-resources/what-is-a-sign-language-interpreter/ (accessed on 23 September 2022).

12. Brooke, J. SUS - a Quick and Dirty Usability Scale. In; 1996; pp. 189–194.

13. Lewis, J. Psychometric Evaluation of the Post-Study System Usability Questionnaire: The PSSUQ.; January 1 1992; Vol. 2, pp. 1259–1263.

14. Hart, S.G. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2006**, *50*, 904–908, doi:10.1177/154193120605000909.

15. Kirakowski, J.; Corbett, M. SUMI: The Software Usability Measurement Inventory. *Br. J. Educ. Technol.* **2006**, *24*, 210–212, doi:10.1111/j.1467-8535.1993.tb00076.x.

16. PIADS – Updates, News and Resources on the Use of PIADS to Measure Impact of Assistive Devices Available online: http://piads.at/ (accessed on 21 September 2022).

17. Team UEQ User Experience Questionnaire (UEQ) Available online: https://www.ueq-online.org/ (accessed on 18 August 2022).

18. Schrepp, M.; Hinderks, A.; Thomaschewski, J. Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios.; June 22 2014; pp. 383–392.

19. Oliveira, T.; Escudeiro, P.; Escudeiro, N.; Rocha, E.; Barbosa, F.M. Automatic Sign Language Translation to Improve Communication. In Proceedings of the 2019 IEEE Global Engineering Education Conference (EDUCON); April 2019; pp. 937–942.

20. ZDGNS Zveza društev gluhih in naglušnih Slovenije Available online: http://zveza-gns.si/ (accessed on 24 August 2022).

21. Alves, A.; Ferreira, S.; Veiga, V.; Silveira, D. Evaluation of Potential Communication Breakdowns in the Interaction of the Deaf in Corporate Information Systems on the Web Http://Www.Sciencedirect.Com/Science/Article/Pii/S1877050912007892. *Procedia Comput. Sci.* **2012**, *14*, 234–244, doi:10.1016/j.procs.2012.10.027.

22. Mohamad, N.; Hashim, N.L. UX Testing for Mobile Learning Applications of Deaf Children. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*, doi:10.14569/IJACSA.2021.0121134.

23. Schnepp, J.; Shiver, B. Improving Deaf Accessibility in Remote Usability Testing. In Proceedings of the The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility - ASSETS ’11; ACM Press: Dundee, Scotland, UK, 2011; p. 255.

24. VINTAR, Š. Lexical Properties of Slovene Sign Language: A Corpus-Based Study. *Sign Lang. Stud.* **2015**, *15*, 182–201.

25. United Nations Convention on the Rights of Persons with Disabilities and Optional Protocol 2006.

26. Oh, J.; Jeon, S.; Kim, B.; Kim, M.; Kang, S.; Kwon, H.; Kim, I.; Song, Y. Avatar-Based Sign Language Interpretations for Weather Forecast and Other TV Programs. *SMPTE Motion Imaging J.* **2017**, *126*, 57–62, doi:10.5594/JMI.2016.2632278.

27. Orero, P.; Martín, C.A.; Zorrilla, M. HBB4ALL: Deployment of HbbTV Services for All. In Proceedings of the 2015 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting; June 2015; pp. 1–4.

28. Schrepp, D.M. User Experience Questionnaire Handbook. 15.

29. Douglas, S.P.; Craig, C.S. Collaborative and Iterative Translation: An Alternative Approach to Back Translation. *J. Int. Mark.* **2007**, *15*, 30–43, doi:10.1509/jimk.15.1.030.

30. RedOrbit HbbTV Emulator Available online: https://chrome.google.com/webstore/detail/redorbit-hbbtv-emulator/mmgfafehampkahlmoahbjcjcmgmkppab?hl=en (accessed on 15 June 2022).

31. HbbTV Specifications Available online: https://www.hbbtv.org/resource-library/specifications/ (accessed on 20 August 2022).

32. Guidelines for Positioning of Sign Language Interpreters in Conferences Including Web-Streaming.Pdf.

33. HBB4ALL Sign Language Interpreting in HBBTV. 18.

34. Sousa, V.; Rojjanasrirat, W. Translation, Adaptation and Validation of Instruments or Scales for Use in Cross-Cultural Health Care Research: A Clear and User-Friendly Guideline. *J. Eval. Clin. Pract.* **2011**, *17*, 268–274, doi:10.1111/j.1365-2753.2010.01434.x.

35. Equal Dreams Placement of Sign Language Interpreters in Media.

36. HBB4ALL Sign Language Interpreting in HbbTV. 18.

37. Quandt, L.C.; Willis, A.; Schwenk, M.; Weeks, K.; Ferster, R. Attitudes Toward Signing Avatars Vary Depending on Hearing Status, Age of Signed Language Acquisition, and Avatar Type. *Front. Psychol.* **2022**, doi:10.3389/fpsyg.2022.730917.

38. DHI 11th DHI Conference in Ljubljana and Zagreb 2021 Available online: https://www.dhi2022slocro.com (accessed on 25 August 2022).

39. Zach T-Distribution Table. *Statology* 2018.

40. The Pennsylvania State University Multivariate Paired Hotelling’s T-Square Available online: https://online.stat.psu.edu/stat505/lesson/7/7.1/7.1.8 (accessed on 23 August 2022).

41. Sign Language Network Guidelines for Positioning of Sign Language Interpreters in Conferences, Including Web-Streaming. **2016**.

42. Orero, P.; Martín, C.A.; Zorrilla, M. HBB4ALL: Deployment of HbbTV Services for All. In Proceedings of the 2015 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting; June 2015; pp. 1–4.

43. EUR-Lex Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the Accessibility of the Websites and Mobile Applications of Public Sector Bodies Available online: https://eur-lex.europa.eu/eli/dir/2016/2102/oj (accessed on 26 August 2022).

44. What Is Directive (EU) 2016/2102? Available online: https://directive2102.eu (accessed on 26 August 2022).