Automatic speech recognition in the booth

Assessment of system performance, interpreters' performances and interactions in the context of numbers

Bart Defrancq and Claudio Fantinuoli
Ghent University | Johannes Gutenberg University Mainz

Automatic Speech Recognition (ASR) has been proposed as a means to enhance state-of-the-art computer-assisted interpreting (CAI) tools and to allow machine-learning techniques to enter the workflow of professional interpreters. In this article, we test the usefulness of real-time transcription with number highlighting of a source speech for simultaneous interpreting using InterpretBank ASR. The system's precision is high (96%) and its latency low enough to fit interpreters' ear–voice span (EVS). We evaluate the potential benefits among first-time users of this technology by applying an error matrix and by investigating the users' subjective perceptions through a questionnaire. The results show that the ASR provision improves overall performance for almost all number types. Interaction with the ASR support is varied and participants consult it for just over half of the stimuli. The study also provides some evidence of the psychological benefits of ASR availability and of overreliance on ASR support.

Keywords: automatic speech recognition, simultaneous interpreting, rendition of numbers, computer-assisted interpreting

1. Introduction

This study reports on a small-scale experiment with in-booth computer-assisted interpreting (CAI). While technological support systems have been widely used in translation for decades, the first modest attempts at technological support for interpreting have only recently been made. The sheer complexity of the interpretation process – with its acoustic input and output, contextual dependence and, in the case of simultaneous interpreting, short parsing window – is one reason why technological support has, until recently, fallen short of interpreters' expectations

https://doi.org/10.1075/target.19166.def
Target issn 0924-1884 | e-issn 1569-9986

(Corpas Pastor and Fern 2016; Fantinuoli 2018). Given the complexity of the interpreting task, it seems clear that interpreters could benefit considerably from technological support during interpretation.

Various technological support options for interpreters may be envisaged, ranging from displaying information in point form on a screen in the booth, to providing interpreters with access to the machine-translation output of an automatic transcription of the source text. However, a support system must deliver accurate information quickly and should not add to the already high cognitive load associated with interpreting. Systems that offer a large amount of information will slow down delivery and place a burden on interpreters' mental processing. In contrast, systems that supply small amounts of targeted information may prove beneficial for interpreters.

Numbers are among the most dreaded source-text features in simultaneous interpreting, and interpreters report them as an important stress factor (Alessandrini 1990). Research shows that accuracy levels for the interpretation of numbers are fairly poor (see Section 2); however, accuracy improves considerably when visual input is made available (Lamberger-Felber 2001; Desmet, Vandierendonck, and Defrancq 2018). Software that is capable of retrieving numbers from source texts and displaying them in numerical form on a screen in the booth could therefore have a significantly positive impact on interpreters' rendition of numbers.

This article reports on the features, usability and use of one such system, namely InterpretBank.[1] The advent of neural Automatic Speech Recognition (ASR) has made it possible to rapidly display reliable transcripts that fit into interpreters' Ear–Voice Span (EVS) and, presumably, could help to improve accuracy rates. In-booth CAI has recently been empirically tested in a number of studies (Prandi 2018), but none have made use of ASR. The model of InterpretBank used in this study, however, displays the output of the ASR as a running transcription in which numbers are presented in (mostly) numerical form and are highlighted. This study, which is the first of its kind, assesses InterpretBank's performance, compares interpreters' performance with and without in-booth support for numbers, and explores their interaction with in-booth support. Section 2 provides an overview of the literature on numbers in interpreting, while Section 3 describes the technology tested during the experiment and how the experiment was set up. The results of the different analyses are grouped in Section 4 and are discussed in Section 5, and the study's conclusions are presented in Section 6.

1. A free online version is available at www.interpretbank.com/asr.

2. Interpreting numbers

Although there is consensus that numbers pose challenges during simultaneous interpreting (Gile 1995; Jones 2002; Setton and Dawrant 2016), requiring specific strategies and triggering seemingly high error rates, there is very little research on them (Mead 2015). Among the studies that have been done, experimental designs dominate: Between 1996 and 2019, at least eight experimental studies have been conducted, involving both professional interpreters and students (Braun and Clarici 1996; Lamberger-Felber 2001; Mazza 2001; Pinochi 2009; Timarová 2012; Korpal 2016; Desmet, Vandierendonck, and Defrancq; Frittella 2019). Only one study is corpus-based: Collard (2019). The results of the experimental studies confirm that numbers are poorly rendered and that error rates typically range between 30% and 70%, with higher error rates among student interpreters (Braun and Clarici 1996; Korpal 2016; Frittella 2019). Such error rates are cause for concern: If, on average, one in two numbers is rendered incorrectly, the profession may face serious reliability issues. However, Collard (2019), the only available corpus-based study, shows that in a sample of more than 700 interpreted numbers in the European Parliament, the error rate only amounts to 21%. The presence of a colleague in the booth, absent in experimental settings, or the availability of documents may explain the discrepancy between Collard's (2019) results and those of the experimental studies. Furthermore, in most experimental studies, error rates may be artificially inflated by overly strict criteria for accuracy. Approximations, for instance, are usually counted as errors, whereas they may be acceptable in several contexts. Given this, it seems clear that a reliable support system in the booth might help interpreters increase their accuracy in terms of rendition.

This is also supported by the fact that error rates drop significantly when interpreters are given documents in the booth (Lamberger-Felber 2001), when they are allowed to take notes while interpreting (Mazza 2001), or when they can see the numbers displayed on a screen in the conference room (Desmet, Vandierendonck, and Defrancq 2018). Lamberger-Felber (2001) and Desmet, Vandierendonck, and Defrancq (2018), for instance, respectively report a decrease in errors of 50% and 70%. Mazza (2001) reports a decrease in errors of about 10% when interpreters are allowed to take notes, compared to when they are not allowed. However, since not all interpreters in Mazza's study took notes when they were allowed to, it is difficult to determine what the effect of note-taking is. Nevertheless, all three studies suggest that the availability of visual numerical input improves interpreters' performance accuracy. The experiment reported in Desmet, Vandierendonck, and Defrancq (2018) makes use of simulated technological support where interpreters are presented with a visual version of the number immediately following its delivery by the source speaker. The 70% drop in error rates measured in this experiment is directly relevant for the purposes of this study, as the experimental conditions are similar.

As far as technological support systems for interpreters are concerned, three requirements should be met. First, since interpreters are likely to spot errors, a fully automatic support system should present very accurate information and should have accuracy levels that far outperform the best-performing interpreters. Systems that do not meet this requirement could negatively affect number renditions, and could reduce the level of trust interpreters place in the system and cause them to abandon the support altogether. Second, the visual input should be presented in an ergonomically suitable format. In Desmet, Vandierendonck, and Defrancq's (2018) experimental design, numbers were displayed on Microsoft PowerPoint slides on a screen behind the speaker. The current number and the two previous ones were visible, and updated as each subsequent number was delivered. Together the three displayed numbers occupied 60% of the slide. The subjects reported that this was satisfactory. However, it is ecologically problematic since conference rooms are unlikely to be equipped with one screen for the speaker and another for the interpreter. For a setup to be ecologically valid, numbers should be displayed in the booth, and in a format that facilitates reading. Third, the support system's latency should be minimal: Interpreters should not have to interrupt their delivery in order to wait for the numbers to be displayed. The crucial 'sound barrier' is assumed to be between one and a half and two seconds, as the reported average EVS of interpreters across several studies is between two and a half and three seconds (Oléron and Nanpon 1965/2002; Christoffels 2004; Defrancq 2015). Crucially, as Collard (2019) shows, interpreters tend to reduce EVS when numbers are delivered. This may have two implications for this study: On the one hand, the use of ASR could relieve interpreters of the burden of reducing EVS in the presence of numbers; however, on the other, if the EVS needs to be reduced independently of the ASR presence, the system's latency may need to be further reduced and well below that of the 'sound barrier' assumed above.

These technical requirements give rise to the following research questions:

1. Does an ASR support system, such as the one implemented in InterpretBank, offer a viable ASR output for interpreters in terms of ergonomics, precision and latency?
2. Does the provision of ASR in booths improve interpreters' performance?
3. How do users experience ASR support and interact with it in the booth?

3. Methodology

3.1 InterpretBank

InterpretBank ASR (Fantinuoli 2017) was used in this experiment. InterpretBank ASR is a prototype of a web-based ASR-supported CAI tool that transcribes, in real-time, speech delivered by a speaker and automatically provides the interpreter with numerals and their unit of measurement, and translation options for terminology (drawn from an event-related terminology database or produced by a machine translation).

The tool's workflow is straightforward. First, the acoustic signal that the interpreter receives in the headset is sent to the sound card of the computer equipped with the ASR-CAI tool. The audio signal is then sent to the InterpretBank Application Programming Interface (API) that operates on a server located in Dresden, Germany, and returns the real-time transcript of the speech. InterpretBank uses the Google Cloud Speech-to-Text API[2] as the ASR of choice. Experimental tests have shown that, compared to its major competitors, the Google Cloud Speech-to-Text API provides the best transcription quality for features useful for CAI integration (Brüsewitz 2019), such as specialized terminology and numbers. Second, the transcription stream is pre-processed, which involves chunking the text stream into units of *n*-words of a fixed size and normalizing them (for example, harmonizing the way numbers are transcribed). Third, for each *n*-words window the units of interest are extracted: Numbers and their units of measurement are detected, and single and multiword grams are looked up in the terminological database loaded in the tool or translated by means of machine translation. In this phase, predicting algorithms can be used to intelligently select the units of interest and increase the usability of the tool (see Vogler, Stewart, and Neubig 2019). Finally, the extracted data are displayed on the computer's monitor.

The InterpretBank ASR prototype has been designed with three different models of data analysis and visualization, thus allowing for different approaches to human–machine interaction in the context of interpreting to be empirically tested.[3] We used Model 1 in our experiment as it was the only available option at the time of the experiment. Model 1 spots numbers and the terminological units

2. https://cloud.google.com/speech-to-text

3. As Model 1 may provide too much visual information for the user (the entire transcript), Model 2 suppresses the text stream and visualizes only the extracted UI in a vertical way (like a TV prompt), with the newest information highlighted on top. Model 3 makes use of advanced algorithms in order to spot the terminology without any background reference (the event glossary) and proposes ad-hoc candidate translations using machine translation as well as the pairs number/unit. The visualization is the same as Model 2.

contained in the event terminology database, and displays the entire transcript with highlighted units of interest. The rationale behind this is to retain the informational context of the units of analysis that may help the interpreter to disambiguate the information (e.g., in terms of co-references). In this model, numbers are displayed as transcribed by the ASR system without performing any kind of normalization (e.g., numbers may be transcribed as digits or as words, depending on the conventions of the particular language).

Since time plays a central role in simultaneous interpreting, in our experiment, effort was made to keep latency to a minimum and within the EVS values reported in the literature (see Section 2). To achieve this, InterpretBank uses the provisional results of speech recognition and not the final results provided by the speech recognition engine. The temporary stream of transcription thus returns provisional results with low latency while the sentence is still unfolding. The transcription is based on a very high approximation of results (i.e., high error rate), redundancies and continuous corrections. In order to use this stream of data in a meaningful way, the ASR engine applies text analyses and transformations through a set of algorithms that mimic the final elaboration of the audio.

3.2 Preliminary test

A preliminary test was carried out with two pre-recorded speeches and no interpreting to see how the system would display the numbers and to test the general setup of the experiment. During the pilot, it became clear that speeches should not exceed five minutes because the experimental version of InterpretBank is set to display the transcription for five and a half minutes without needing to reactivate the service. Very high delivery rates cause the system to overload, increasing the probability of missing numbers and decreasing precision. It was therefore decided to limit the speeches to approximately 650 words and to limit the delivery rate. The speaker managed to keep delivery rates within a range of 105 to 122 words per minute for the experiment. New speeches were drafted for the experimental phase.

ASR output is displayed in a text field in the bottom half of the screen (see Figure 1). Numbers appear in a bigger font and in red. InterpretBank does not automatically scroll down when the text field is full, and manual scrolling is needed after a certain point. Some numbers were shown orthographically (i.e., 'two', 'three' and 'ten'), which was systematic for the number '1'. It was therefore decided not to take the number '1' in consideration. 'Million' and 'billion' were also always shown orthographically, even when combined with another number, and were not highlighted. It was nevertheless decided to include that type of number in the experimental phase, because removing them would have severely limited the relevance of the study.

3.3 Equipment

The preliminary test and the experiment were run in a Televic Education AVIDAnet® Smart Interpreter Lab. The lab consists of a conference table and ten booths. The conference table is equipped with a trainer PC and two 17″ screens, one of which is used for system and booth management and the other for regular PC applications. The latter was used to monitor the ASR output while the source speech was presented.

Six booths were used for the experiment, all of which allow for visual contact with the speaker in the conference room. Each booth is equipped with two interpreter consoles, a 17″ screen and a webcam. The screen is partially inserted in the booth table and inclined at an angle of 30–35°. Interpreters can switch between three displays on the screen: trainer's PC screen, table camera image and booth PC screen. For the purpose of the experiment, the screens in Booths 2, 3 and 4 were connected to the trainer's PC screen displaying the ASR output, as shown in Figure 1. The screens in Booths 7, 8 and 9 displayed the camera image of the speaker.

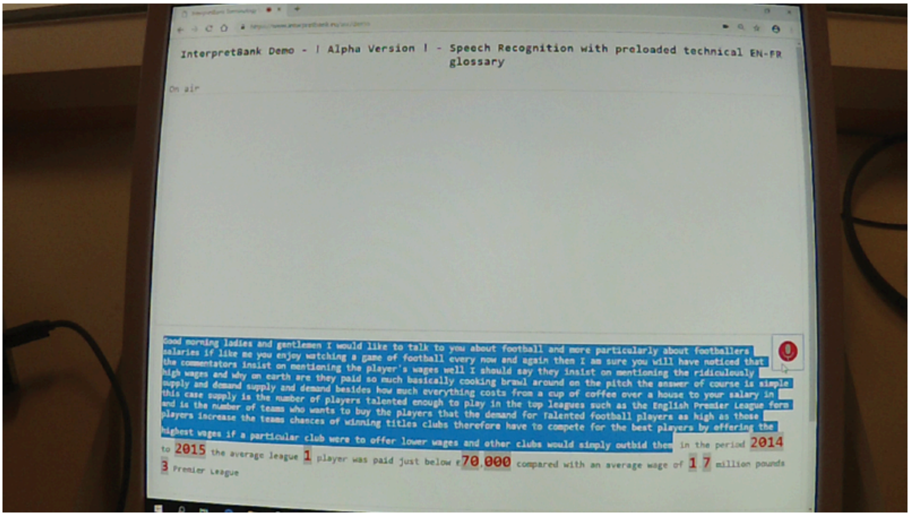

Figure 1. Booth display of ASR and numbers

The AVIDAnet® Smart Interpreter Lab allows for the synchronized recording of the source speech (audio), the source speaker (video), the interpreters' performance (audio) and their behavior in the booth (video). For the experiment, video recording of the source speaker was not required and was therefore switched off. Video recording in the booth was crucial to find out whether participants sought support from the ASR. As the screen is situated well below the gaze line for the speaker, it was possible to visually identify the instances when the participants consulted the screen. Unfortunately, in one booth the webcam was poorly oriented and the video could not be used. Observing a change in gaze does not imply that the interpreter actually saw the number and used it to interpret; we therefore refer to these observations as 'presumed' use of ASR.

One additional booth was used to record the screen display of the ASR output in the booth and the audio input from the interpreter's console (source speech). This was done to determine the ASR's precision and measure its latency. A Zoom Q2n camera was used for this purpose.

3.4 Participants

The participants were regular, full-time students enrolled in the 2019 postgraduate program in conference interpreting at Ghent University. All six participants were female and aged between twenty-three and twenty-four. The participants' A-language is Dutch, which was the target language of the experiment. The source language of the speeches was English. Three of the six participants have English as a B-language and three as a C-language. All participants have a master's diploma in interpreting focusing on modes and techniques for public service interpreting. They were acquainted with consecutive and simultaneous interpreting, and had sixty hours of simultaneous interpreting experience from English into Dutch

The experiment took place on 17 May 2019. The participants did not receive specific training on the use of ASR in the booth; however, they had had ample opportunity to train with other types of information displayed on the screens, such as text and PowerPoint presentations. They were informed that the experiment dealt with the interpretation of numbers and that they would be provided with the output of an automatic transcription tool to help them with the task. Admittedly, this may have caused them to focus more on the rendition of numbers during the experiment than they normally would have. However, for the purpose of testing and assessing the ASR prototype, we had little choice but to inform them right from the start about the experimental setup.

Our focus was on ecological validity and we tried to mimic real training conditions as usually implemented in the Ghent program. For this reason, we chose not to use pre-recorded speeches, and instead asked one of the trainers to deliver the speeches live. This decision limited the number of participants, as speeches cannot be delivered multiple times in completely identical ways, which in turn limits the generalizability of the results. We also sought to keep the population as homogeneous as possible in terms of experience with simultaneous interpreting and therefore decided to only invite students of the Ghent program in conference interpreting to participate.

3.5 Speeches

Four English speeches were prepared (and delivered) by an interpreter trainer who is near-native in English. She was asked to prepare speeches of approximately five and a half minutes for training in simultaneous interpreting on topics of her choice and with a clear rhetorical structure. The speeches included an introduction of approximately one minute without numbers, followed by a descriptive part that included numbers and ended with a conclusion. The trainer was asked to include at least twenty numbers of various types and degrees of complexity per text but was not given a maximum number. Among the number types, we sought to represent positive integers, decimal numbers and dates. Tables 1, 2 and 3 provide detailed information on the properties of the speeches and the numbers included in them.

Table 1 shows that three of the speeches exceeded the five-and-a-half-minute limit. Source speech delivery rates varied between 105 and 122 words per minute, very close to the ideal speech rate for interpreters, and close to the range (100–110 words per minute) at which trainees in Korpal and Stachowiak-Szymczak (2020) performed best.

The difficulty levels of the source speeches, which were fairly similar, were measured post hoc so that this could be taken into account in interpreting the results. On the Flesch Reading Ease Index, the difference between the most difficult and the easiest speech is less than 10 points (or two grades in school). On the Gunning Fog Index, the difference is less than 2.5 points (or two and a half grades in school). According to both indexes, Speech 2 is the easiest. There is no agreement between the two indexes on the ranking of the other speeches. This demonstrates that, with the exception of Speech 2, the difficulty levels of the speeches were comparable.

In total, 119 numbers were acoustically presented to the participants (see Table 2). Speech 3 contained nearly twice as many numbers as each of the other three speeches, which was mainly due to the fact that the speech drew systematic comparisons between two Belgian towns. This potentially affects the comparability of the results across speeches and across groups.

Table 1. Properties of source speeches

Speech number and topic	Length (words)	Duration (mins)	Speech rate (words/min)	Flesch reading ease index	Gunning fog index
1. Footballers' wages	628	5'58"	105	57.73	11.99
2. Real estate agents	671	5'32"	121	62.60	10.78
3. Aalst carnival and Ghent festival	715	5'55"	121	59.08	13.11
4. Amazon	657	5'24"	122	54.82	12.22
Total	**2671**	**22'47"**			

Table 2. Frequencies of numbers per number type

	Number type			
Speech number and topic	**Integers**	**Decimal numbers**	**Dates**	**Total**
1. Footballers' wages	17	2	6	25
2. Real estate agents	16	3	2	21
3. Aalst carnival and Ghent festival	31	5	9	45
4. Amazon	12	4	12	28
Total	**76**	**14**	**29**	**119**

A Fisher's Exact test confirmed that although the different number types are not identically distributed across the texts, their distribution does not differ significantly ($p = 0.15$). Differences in the participants' performance can therefore not be attributed to differences in the frequency of particular number types.

Numerical complexity varies across languages: In English numbers like '100' and '1000' consist of two numeric items ('one hundred', 'one thousand'), whereas in many other languages, such as the target language in this study, Dutch, they only consist of one item (*honderd, duizend*). Only the complexity of the input language was taken into consideration, and we distinguished four levels of complexity according to the number of numerical items included in the spoken form of the number:

- Level 1 refers to numbers with one or two numeric items (e.g., '2', '63' or '40.5').
- Level 2 refers to numbers with three or four numeric items (e.g., '124', '1024', '310 000' or '7.6 million').
- Level 3 refers to numbers with five or six numeric items (e.g., '1130', '1 406 000').
- Level 4 refers to numbers with more than six numeric items (e.g., '17 345 133').

Table 3 provides an overview of the frequency of numbers according to level of complexity in each speech.

Table 3. Frequencies of numbers according to level of complexity

	Complexity level				
Speech	**Level 1**	**Level 2**	**Level 3**	**Level 4**	**Total**
1	8	10	5	2	25
2	7	11	2	1	21
3	17	23	5	0	45
4	7	17	3	1	28
Total	**39**	**61**	**15**	**4**	**119**

A Fisher's Exact test confirmed that although the different complexity levels are not identically distributed across the texts, their distribution does not differ significantly ($p = 0.62$), and therefore differences in the participants' performance cannot be attributed to the unequal distribution.

3.6 Procedure

The participants were not informed in advance about the topics of the speeches. Participants were divided into two groups of three, with each group comprising at least one student with English as a B-language and one with English as a C-language. Both groups interpreted alternately in booths with and without ASR support (see Table 4), with five minute breaks between each speech. In all, the experiment lasted forty-five minutes.

At one point during Speech 1, the manual scroll down failed and two numbers remained invisible to the participants. These numbers were excluded from the analyses of interpreter performance. Speeches 1, 2 and 3, which exceeded five and a half minutes, were not fully transcribed by the ASR. As a result, fourteen numbers were not displayed on the screen. These numbers were not disregarded in the analysis as they provided an unanticipated opportunity to study the effect of the sudden loss of ASR support. Tables 5 and 6 show the number of items that were actually displayed by the ASR during the experiment.

The participants were asked to fill out a questionnaire comprising six main questions. The questions enquired into their use of the ASR, their perceived usability of the tool, their assessment of the tool's accuracy, the extent to which the tool interfered negatively with the interpreting task, and their preference for a display format (i.e., only numbers are displayed), with a last question asking

Table 4. Group composition and counter-balancing of ASR support

Group member	Speech			
	1. Footballers' wages	**2. Real estate agents**	**3. Aalst carnival and Ghent festival**	**4. Amazon**
S1.1	No support	Support	No support	Support
S1.2	No support	Support	No support	Support
S1.3	No support	Support	No support	Support
S2.1	Support	No support	Support	No support
S2.2	Support	No support	Support	No support
S2.3	Support	No support	Support	No support

Table 5. Numbers displayed in relation to numbers presented acoustically (in brackets), per number type

	Number types			
Speech	**Integers**	**Decimal numbers**	**Dates**	**Total**
1	11 (17)	1 (2)	6 (6)	18 (25)
2	15 (16)	3 (3)	2 (2)	20 (21)
3	23 (31)	5 (5)	9 (9)	37 (45)
4	12 (12)	4 (4)	12 (12)	28 (28)
Total	**61 (76)**	**13 (14)**	**29 (29)**	**103 (119)**

Table 6. Numbers displayed in relation to numbers presented acoustically (in brackets), per level of complexity

	Complexity level				
Speech	**Level 1**	**Level 2**	**Level 3**	**Level 4**	**Total (displayed)**
1	5 (8)	9 (10)	4 (5)	0 (2)	18 (25)
2	7 (7)	10 (11)	2 (2)	1 (1)	20 (21)
3	10 (11)	22 (23)	5 (5)	0 (0)	37 (45)
4	7 (7)	17 (17)	3 (3)	1 (1)	28 (28)
Total	**29 (39)**	**58 (61)**	**14 (15)**	**2 (4)**	**103 (119)**

for additional comments they might have. The first four questions were answered using a five-point Likert scale (never – seldom – some of the time – mostly –always). The fifth question (which asked whether participants would prefer a format in which only numbers are displayed) could be answered with either 'Yes' or 'No'. The sixth question was an open question. Questions 1, 3 and 4 had open-ended follow-up questions on the type of information the participants had drawn from the ASR (only numbers or other items in the transcript), their reaction to perceived errors in the ASR output, and their interactions with the running transcript.

3.7 Data processing and analysis

The audiovisual recordings of the screen display in Booth 1 were analyzed with REAPER (https://www.reaper.fm/). Time tags were placed at four points: (1) at the onset of the acoustic signal corresponding to the number; (2) on the first video frame in which a part of or the whole number was displayed in numerical form; (3) at the end of the number's acoustic signal; and (4) on the first video frame that displayed the final version of the number in numerical form. Whether the transcribed number was correct was irrelevant. Due to InterpretBank ASR's workflow properties (see Section 3), the transcript changed shape considerably. For example, one instance of 300 000 went through the following stages (Example (1)):

(1) three > 3 > 300 > 300000 > 300,000

To illustrate how time tags were added, in Example (1) we tagged the second stage (onset of the numerical display) and the fourth stage (complete version except for punctuation). We excluded the frames that showed the orthographic version of the number and we considered the addition of the comma during the last stage irrelevant for the human recognition of the number. Exceptions were made in the case of numbers that only displayed orthographically, such as 'two', 'million' and 'billion'. When the punctuation was considered relevant – as for instance in decimal numbers – the final time tag was placed when the full decimal form was displayed.

Latency for each number was determined by subtracting the onset and final tags of the video recording in Booth 1 (output of ASR) from the corresponding time tags added to the articulation of the number. Although REAPER displays time in milliseconds, it was decided to round off to centiseconds as the video recording was carried out at a rate of sixty frames per second.

The participants' interpretations were recorded (as well as the source speech) and manually checked for number accuracy. Performance was only assessed in terms of the accuracy of number rendition, since other criteria are beyond the scope of this article. The assessment was conducted at two levels, in line with Collard (2019): First, a general distinction was made between renditions and omissions; and second, renditions were classified into different types (see Table 7).

Table 7. Categories of rendition

Category	Explanation
Complete rendition	The number is correctly and completely rendered, occasionally after a first erroneous or incomplete attempt. This includes cases where a year is rendered in short form (e.g., '1997' rendered as '97').
Approximation	The order of magnitude is correct, but the number is rounded off (e.g., '1864' rendered as '1800').
Related substitution	The number is replaced by another number which bears some resemblance to the original number. The relationship can be phonological (e.g., '14' rendered as '40') or syntactic (e.g., '47' rendered as '470' or as '74').
Unrelated substitution	The whole number or some parts of the number are replaced by a number with no resemblance. Substitution can be partial (e.g., '72' rendered as '73') or complete (e.g., '58' rendered as '140').

There is an obvious accuracy cline across these types that can be described as follows: Complete rendition > approximation > related substitution > unrelated substitution > omission. We classified instances in which rendition types were combined as examples of types that are lowest on the accuracy cline. For instance, when the source number '69 381' was rendered with 'more than 6 000', it was analysed as a case of related substitution (in which the relationship was syntactic because the order of magnitude is shifted) and not as a case of approximation. Since the accuracy cline is only used in the analysis to aid decisions for hybrid cases, the discussion of accuracy is not pursued here.

The participants' performance was analysed with simple inferential statistics, such as a chi-squared test. The use of inferential statistics for so small a population may seem controversial, but as we are mainly concerned with within-person and within-item variation across support conditions, inferential statistics are informative. The significance threshold was set at 0.05.

4. Results

4.1 Findings regarding ASR

4.1.1 *Ergonomics*

With regard to the participants' general assessment of the usability of the ASR support system, four participants indicated that the system is 'sometimes' usable, while two felt that it is 'often' usable. Interestingly, in the open comments, two participants described the usability of the system in terms of metaphors such as 'safety net' and 'emergency backup'.

Four participants reported that they were sometimes or often distracted by the running transcription; three of whom claimed that they had made mistakes due to the transcription. Three participants indicated that it would be better if only numbers were displayed, while three indicated a preference for numbers and units. Four participants indicated that they did use the running transcript for other items, such as names.

Further comments related to the way numbers were displayed. Two participants felt that numbers should be displayed in their final version only, because they were distracted by the changing shapes. This is surprising since the latency measurements (see Section 4.1.2) show that InterpretBank displayed the final version of all numbers at latencies below students' average EVS. The comments thus seem to indicate that the participants in question either have a very short EVS or that they consulted the ASR output as soon as they heard the number (and without having reached the number in their own delivery).

Two participants indicated that they would prefer the numbers to be displayed differently, either in a separate frame of the screen or in a bigger format. Crucially, only one participant reported that she stopped using the support system during one of the interpretations. It thus seems that, despite its shortcomings, the participants appreciate the benefits of the ASR support, especially as a tool that provides a safety net.

4.1.2 *Latency*

The average latencies and ranges for the numbers are presented in Table 8.

The results vary considerably across the speeches. One fairly obvious reason is related to delivery rate, which was slower in the case of Speech 1 (105 words per minute) compared to the other speeches (121–122 words per minute). The differences between Speeches 2 to 4 were not analyzed in detail, but it seems that short numbers and dates trigger slightly longer latencies: Speech 4 contains twelve dates of a total of twenty-eight numbers, whereas Speech 1 only contains six dates of a

Table 8. Average ASR latencies and ranges for numbers in centiseconds (cs)

Speech	Onset latency		End latency	
	Average (cs)	Range (cs)	Average (cs)	Range (cs)
1	0.84	0.54–1.12	0.28	0.05–0.54
2	0.97	0.73–1.41	0.66	0.38–1.19
3	1.19	0.82–2.23	0.64	0.21–1.45
4	1.63	0.75–2.56	1.04	0.44–1.78
All speeches	1.20	0.54–0.256	0.69	0.05–1.78

total of twenty numbers. Mispronunciations and self-repairs also have a limited effect.

It appears that InterpretBank catches up with the speaker as the process unfolds: Onset latencies were, on average, half a second higher than end latencies. Furthermore, end latencies all fell below the average EVS reported in the literature (see Section 2). Only a handful of cases broke the 'sound barrier' of one and a half seconds. This means that, provided interpreters maintain an average EVS, the number is readable in its final version before interpreters reach the point at which they would deliver it. InterpretBank thus seems to offer a usable ASR for the booth.

4.1.3 *ASR precision for numbers*

The precision figures shown in Table 9 are for numbers that were displayed, and exclude numbers that were not displayed, which typically occurred at the end of the speeches when the ASR stopped transcribing. Two numbers in Speech 1 were transcribed but were not displayed in time due to the failure of the automatic scroll down. These are nevertheless included in the data, causing a small discrepancy with totals in Tables 5 and 6.

Table 9. ASR precision in terms of numbers

Speech	Number of displayed numbers	Number of correctly displayed numbers	Precision (%)
1	20	20	100
2	20	17	85
3	37	37	100
4	28	27	96
All speeches	105	101	96

With 96% of the numbers displayed correctly, the ASR's precision is in line with the results of similar studies (see Brüsewitz 2019). Errors occurred due to the speaker's self-repairs ('44 500' for 'four thou/ four thousand five hundred' in Speech 2; '1 and 160 000' for 'one hu/ one hundred and six thousand' in Speech 4), to partly orthographic transcription ('2000 and 2' for 2002) and to homonymy ('to' instead of '2'). In all cases, the ASR's precision is higher than interpreters' accuracy levels reported in experimental and corpus-based research. ASR therefore has the potential to help interpreters improve their accuracy.

Five participants indicated that they 'sometimes' or 'seldom' spotted errors in the output of the support system; only one participant indicated that she 'never' spotted an error. Interestingly, two participants claimed that they did not make an error when the ASR displayed an incorrect number. As will be shown in Section 4.3.1, in one case, all participants provided with the incorrect ASR output made an error.

4.2 *Findings regarding presumed use of the ASR*

The findings on the participants' presumed use of ASR are based on the results of four participants. Participants sought ASR support in 55% of the cases, as can be seen in Table 10. There was considerable variation across participants and speeches.

Table 10. Breakdown of presumed use of ASR support

	S2.2		S2.3		S1.2		S1.3	
Speech	**Support not sought**	**Support sought**	**Support not sought**	**Support sought**	**Support not sought**	**Support sought**	**Support not sought**	**Support sought**
1	5	13	5	13	–	–	–	–
2	–	–	–	–	5	15	13	7
3	32	5	7	30	–	–	–	–
4	–	–	–	–	14	14	11	17
Total	**37**	**18**	**12**	**43**	**19**	**29**	**24**	**24**
	Support not sought				**Support sought**			
Total	92 (44.7%)				114 (55.3%)			

Participants S2.2 and S1.2 sought ASR support less often for the second speech than for the first. S1.3 followed an opposite trend. The low consultation rate can therefore not be attributed to participants' lack of familiarity with ASR in the booth: It may be assumed that the participants were more familiar with ASR after the first speech than before, yet the overall consultation rate dropped from the first speech to the second. This suggests that at least two participants felt confident enough in their own skills to be able to cope with most numbers without ASR support.

Finally, we also sought to determine whether the extent to which participants seek ASR support varies according to number type, assuming that more complex numbers, including decimals, are more likely to encourage interpreters to seek support. Surprisingly, this turns out not to be the case, as shown in Figure 2. The proportion of numbers for which support is sought or not sought remains stable across complexity levels. Support is sought slightly less often in the case of dates, but the difference is not significant (see Table 11). In contrast, participants appear to seek ASR support for decimals.

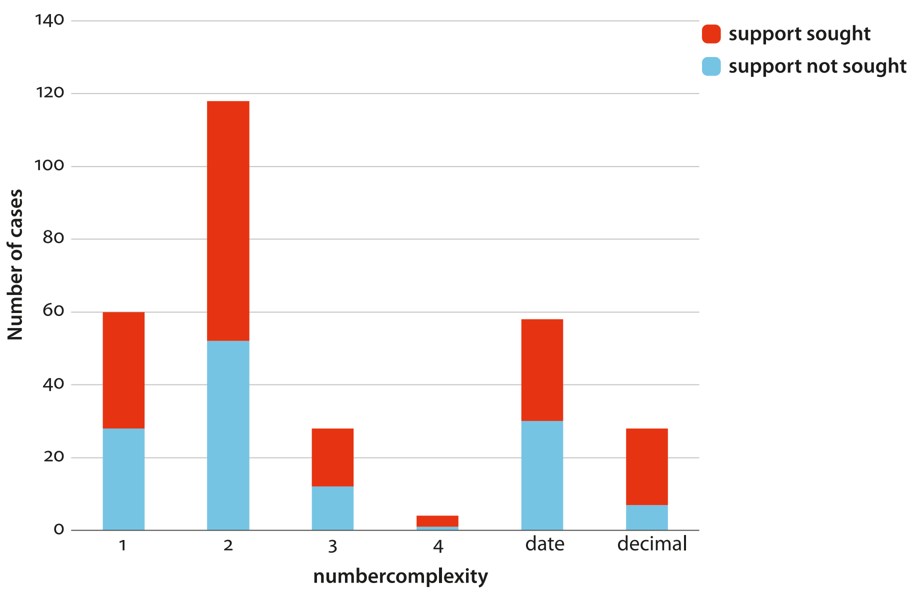

Figure 2. Use of ASR across different number types and complexity levels

Table 11. Statistical data on frequency of use of ASR across different number types

	χ^2	df	p
Complexity levels 1–4	0.766	3	0.87
Dates	1.797	1	0.18
Decimal numbers	4.870	1	0.02

The results in Table 11 could reflect the participants' lack of familiarity with ASR support in the booth. The participants were not specifically trained to use the system and seem unaware of when it could be most useful. Several participants commented on the questionnaire that one needs to get accustomed to the system or that training in its use is needed. Alternatively, it is also possible that when interpreters experience high cognitive load caused by complex numbers, the incentive to use ASR is offset by the anticipated extra load involved in the consultation of written information displayed on a screen.

4.3 *Findings regarding performance*

4.3.1 *Renditions, ASR availability and presumed use*

Number renditions appeared not to be influenced by the difficulty of the speech. Both readability indexes indicated that Speech 2 was the easiest, but a chi-square test of the association between speech and renditions did not yield significant results ($\chi^2 = 16.234$; $df = 12$; $p = 0.18$). Figure 3 shows the frequencies of the different rendition types across participants in contexts with and without ASR support. The total number of renditions is not balanced across the conditions as some renditions made in contexts with ASR support involve cases where the ASR was not displayed due to technical limitations of the ASR prototype. The availability of ASR support clearly increases the share of complete renditions in the total number of renditions (from 67.7% to 90.2%) and drastically reduces the number of omissions (from 15.8% to 3.5%). This means that participants' accuracy improves by nearly a third (a 22.5% gain on 67.7%) with ASR support, which is lower than the gain reported in Desmet, Vandierendonck, and Defrancq (2018).

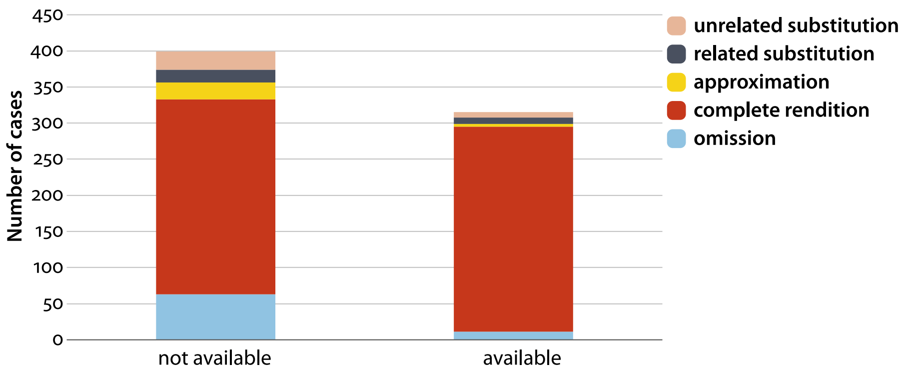

Figure 3. Rendition types with and without ASR support

The chi-square test ($\chi^2 = 54.258$; $df = 4$; $p < 0.001$) confirms that the availability of ASR support is significantly associated with the frequencies of rendition types. ASR support increases the share of complete renditions and reduces the share of omissions and all other rendition types.

To check whether the loss of ASR support influences the participants' renditions, we extracted the renditions for interpretations in ASR-supported booths where the ASR had been temporarily unavailable. In total, fourteen such instances occurred yielding a total number of forty-two renditions. Accuracy levels seem to plummet in such cases: Complete renditions account for only 50% of the renditions. Interestingly, when ASR becomes unavailable, accuracy levels fall below 69.1%, the level recorded in booths where ASR had not been available throughout the speech. It should be noted that this condition was not counterbalanced: Thirteen of the fourteen numbers were presented to the same group in the same condition (Group 1 with ASR support and Group 2 without). Nevertheless, the results seem to indicate that the participants tended to over-rely on ASR once it was offered and struggled to recover when it was suddenly withdrawn.

It should be stressed that Figure 3 reports on the difference between the availability and non-availability of ASR support, and does not take into account whether the participants actually consulted the ASR that was available. We therefore also analyzed the gaze data from the booth cams to investigate rendition types against gaze orientation as evidence of seeking ASR support. Figure 4 presents the results of that analysis. It should be noted that the data in Figure 4 only cover data from two booths. As already mentioned, ASR support was offered in three booths, but the camera was ill-positioned in one.

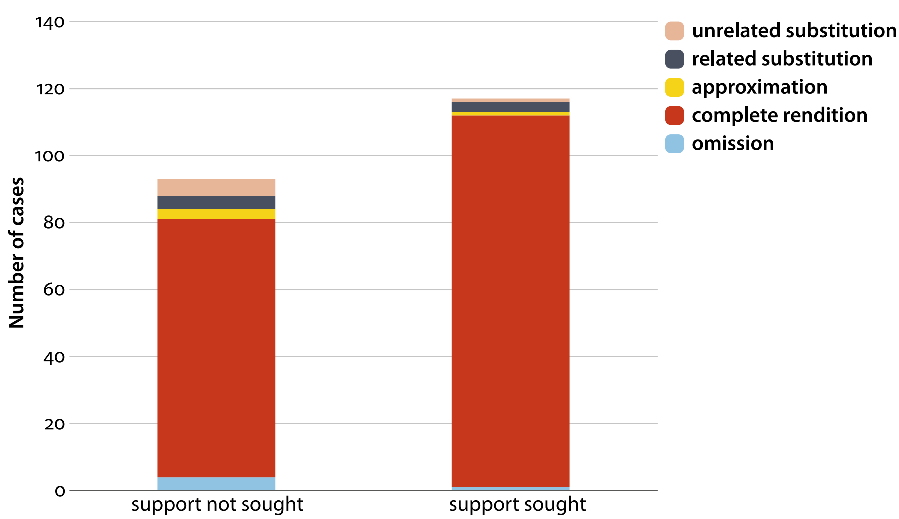

Figure 4. Rendition types and use of ASR support

The share of complete renditions increases to 94.9% when ASR support is sought, compared to 82.8% when no support is sought. With the exception of related substitutions (which are similar in both conditions), the remaining three rendition types' shares are reduced when ASR support is sought. As less data was available in this case, a Fisher's Exact test was performed, which reveals that attempts by the participants to seek ASR support is significantly associated with the distribution of rendition types (Fisher's Exact = 8.740; $p = 0.03$).

Intriguingly, it appears that the participants' accuracy improves when they are offered ASR support, whether they consult it or not. Complete renditions account for 67.7% of renditions when no ASR support is available, compared to 82.8% when ASR is available but not sought. One explanation for this may be that when ASR support is available, participants are less likely to consult it for less difficult numbers, which they are likely to render correctly. However, as noted in Section 4.2, this is not the case: Participants sought ASR support to a very similar extent, irrespective of the types of numbers. At this stage, we can only speculate that the availability of support could have a psychological effect, reducing stress and/or boosting confidence, and that this may lead to the more accurate interpreting of numbers.

Also worth noting is the fact that in three out of six cases where participants sought support and did not deliver a complete rendition, the ASR displayed an inaccurate transcription of the number. This was the case for '106 000', which was transcribed as '1 and 160 000' following a self-repair by the speaker. In this case, both participants sought ASR support and delivered '160 000'. This may also be evidence of an over-reliance on technology.

4.3.2 *Results per interpreter*

The data were broken down per participant to investigate whether ASR availability was beneficial for all participants. The data in Figure 5 refer only to ASR availability and do not take into account whether participants used the ASR support. In the case of the latter, the data are too sparse to allow for statistical testing. Figure 5 shows that the share of complete renditions increases for five of the six participants when ASR is available. Gains in accuracy range between 11.5% (S2.2) and 44.2% (S2.1). For S1.2, complete renditions decline by approximately 6%.

However, it should be noted that for four participants, S1.1, S1.2, S1.3 and S2.2, the difference between their performance with and without support does not reach significance. For S1.1 and S1.3, the difference is near-significant (see Table 12).

Two considerations may help to make sense of these patterns: One the one hand, S1.2 and S2.2 have the highest accuracy levels when they are not offered ASR support (with complete renditions accounting for 81.7% and 82.3% of their

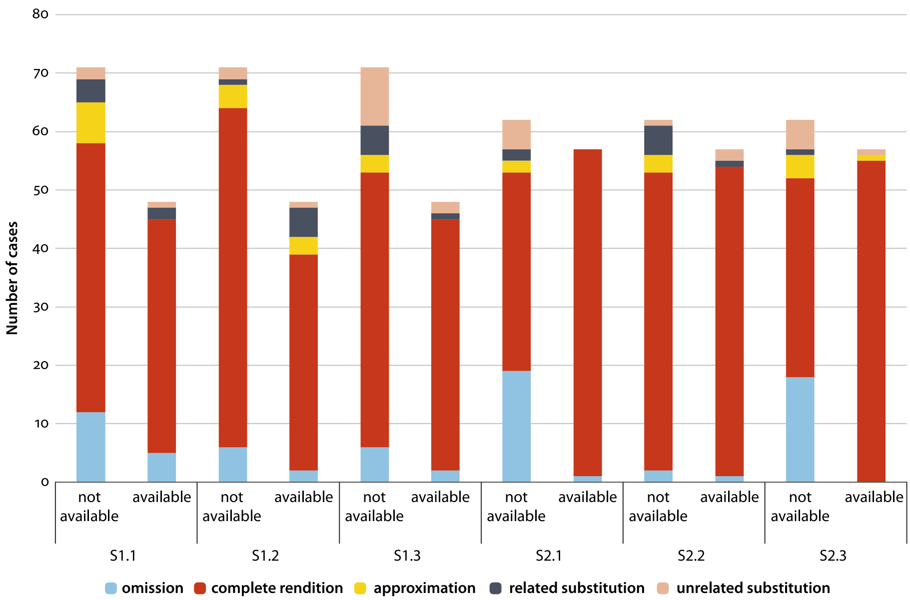

Figure 5. Rendition types with and without ASR support per interpreter

Table 12. Results of the Fisher's Exact test for rendition types per interpreter

Participant	Fisher's Exact	*p*
S1.1	7.330	0.09
S1.2	5.342	0.25
S1.3	8.208	0.05
S2.1	31.585	< 0.001
S2.2	5.760	0.25
S2.3	31.434	< 0.001

renditions, respectively). It is therefore not surprising that the presence of ASR support makes very little difference for participants S1.2 and S2.2. Furthermore, S1.2 and S2.2 sought ASR support less often the second time they were offered it, compared to the first. This suggests that they may have felt that ASR was of little help to them.

4.3.3 *Renditions and number type*

We investigated whether the availability of ASR support was associated with particular distributions of rendition types. The data only reflect availability, as the dataset related to the actual seeking of support is too limited for statistical testing. Figure 6 shows the results for the four levels of numeric complexity.

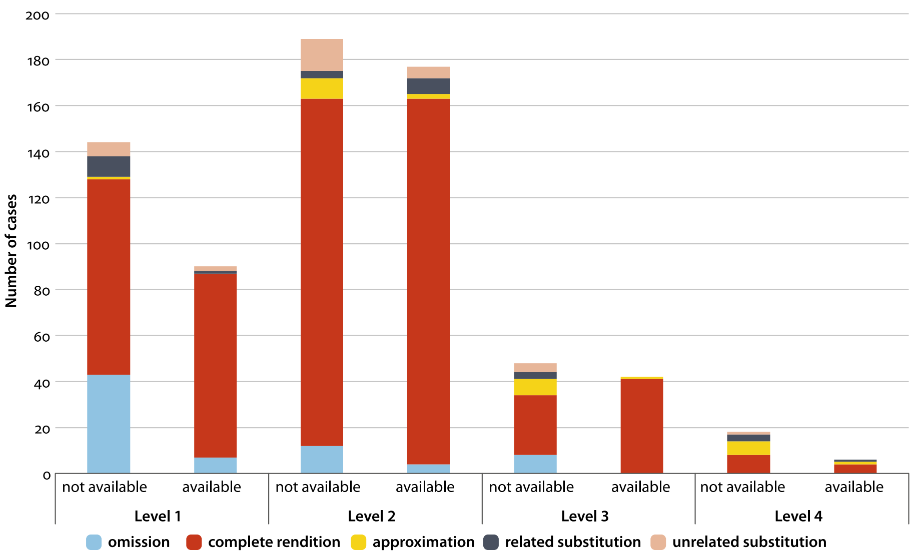

Figure 6. Rendition types with and without ASR support per level of complexity

ASR availability is significantly associated with the distribution of rendition types (see Table 13), except for the most complex number types (i.e., Level 4). This may be due to the sparsity of the data available for that level.

Table 13. Fisher's Exact test for rendition types with and without ASR support per number type

	Fisher's Exact	*p*
Level 1	25.296	< 0.001
Level 2	13.901	< 0.01
Level 3	22.236	< 0.001
Level 4	1.416	0.11

For all complexity levels, the availability of ASR support is associated with a higher number of complete renditions. The gains in accuracy are highest for numbers at Level 3 (numbers consisting of five or six numeric units), where the share jumps from 54.2% without ASR support to 97.6% with ASR support (an increase of just over 80%). For numbers at Levels 1 and 4, complete renditions increase by around 50% and for Level 2, by only 12.4%. The sharp increase for

Level 1 is surprising as the numbers are not complex and probably impose the least cognitive load on interpreters and therefore have a better chance of being interpreted accurately without support. However, as Figure 6 shows, omission occurs frequently if no ASR support is available, probably because the items are acoustically very short and are more likely to go unnoticed. It is worth noting that for both conditions (i.e., interpreting with and without ASR support), the number of omissions decreases as number complexity increases.

The same analyses were carried out for decimal numbers and dates. The data in Figure 7 show an increase of complete renditions in both cases when ASR is available, but only a modest increase for dates, where accurate renderings are also very frequent in the unsupported condition. For decimal numbers, the share of complete renditions increases by 65% (from 54.8% to 90.5%). Despite the availability of ASR, the participants tended to deliver slightly more unrelated renditions in the case of decimal numbers.

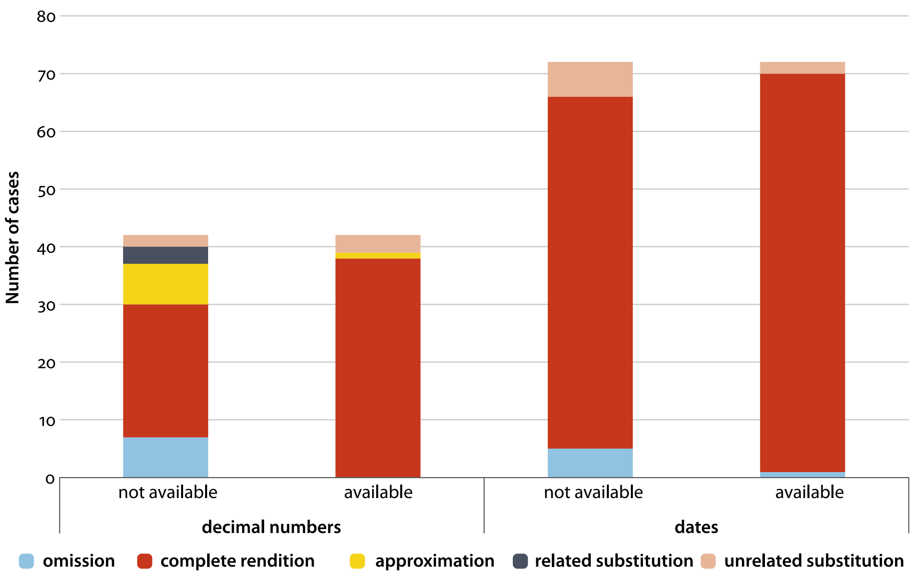

Figure 7. Rendition types with and without ASR support per number type

The significance of these differences, determined using a Fisher's Exact text, is shown in Table 14. ASR availability is significantly associated with the frequency of rendition types in the case of decimal numbers, but not in the case of dates, where the association approaches significance.[4]

4. As suggested before, dates are easier to render accurately than other number types. In our data, irrespective of the availability of ASR support, complete renditions of dates are indeed sig-

Table 14. Fisher's Exact test for rendition types with and without ASR support per number type

	Fisher's Exact	*p*
Decimal numbers	18.248	< 0.001
Dates	4.875	0.06

The availability of ASR support thus positively affects the rendition of almost all number types. For dates and very complex numbers, there is a more modest increase of complete renditions compared to other number types.

5. Discussion

The results of the study provide fairly unambiguous answers to the research questions, but also raise new questions. The answers concern the system's relevance for interpreters: InterpretBank does offer viable ASR support in the booths which helped most participants improve the accuracy with which they interpret numbers. This confirms Desmet, Vandierendonck, and Defrancq's (2018) findings about the usefulness of technological support that displays numbers on a screen and previous work showing that the availability of visual numerical input increases the accuracy of number renditions (Lamberger-Felber 2001). Since the participants in this study were students, it remains to be seen whether professional interpreters benefit to the same extent or not.

The other questions concern the participants' interaction with the technology. First, as pointed out in Section 4, we found that the participants seek ASR support parsimoniously and do not seem to prioritize on the basis of the numbers' features. Several factors may contribute to this. On the one hand, the participants involved in the experiment were not trained to use the system and clearly lacked familiarity with it. A logical follow-up study to this one would therefore need to study the effect of providing specific training for interpreters. On the other hand, there is the issue of the extra cognitive load induced by the use of ASR. ASR output is an extra source of information which requires the allocation of attentional resources. The higher the cognitive load of interpreters, the less likely they will be to allocate their attentional resources to external sources. This seems to be exacerbated by the way the ASR output is displayed, as the participants indicated that they were distracted by the running transcript. This ultimately means that interpreters are least likely

nificantly more frequent than complete renditions of numbers with the same degree of complexity (i.e., numbers at Level 2).

to use ASR when they need it most (i.e., in the case of very complex numbers). In this respect, the experiment could be replicated changing the visualization feature and showing only the numbers without the running script. It is, in fact, reasonable to assume that reducing the visual input to the bare minimum could increase the usability of the technology.

Second, the participants sought support to different extents. On the one hand, their presumed use seems to be correlated with their performance in interpreting numbers without ASR support: The better they are at the task in general, the less they are inclined to consult the ASR, especially after a first experience with this support. This may be related to the fact that such participants also gained less benefit from the support (and in one case even suffered). The question therefore arises whether the high-performing participants became aware of the limited benefits offered by the ASR and consequently tended to ignore the support. If this were confirmed, it would raise crucial concerns about the future of ASR in the interpreting setting. Potential professional users who can be presumed to attain high accuracy rates for numbers could be put off by a perceived lack of personal benefit.

Third, the distinction that was made in this study between the availability of ASR support and its presumed use brought another intriguing aspect of human–technology interaction to light: The mere availability of ASR already improved the participants' accuracy, irrespective of whether it was used. This finding cannot be explained by some natural tendency to abstain from support for 'easy' numbers, as it was shown that there was no meaningful relationship between number types and rates of presumed ASR use. We hypothesized that the availability of support might reassure interpreters and reduce stress, leading to better performance. More research is needed to confirm this. If confirmed, it would mean that the contribution of ASR to cognitive load is even more complex than expected. On the one hand, as an additional source of information, ASR drains attentional resources, leaving less available for the interpreting task. On the other hand, ASR availability seems to reduce stress, which is likely to lead to a more efficient allocation of resources.

Finally, the study also yielded some evidence of over-reliance on ASR. When the participants were faced with the sudden loss of the ASR, their performance dropped below the average levels of interpreting without the support. The same happens when the system offers an incorrect transcription of the number. The instances in which this occurred were not frequent enough to fully explore the risks ASR poses. The questionnaire also brought to light that the participants did not assess themselves well enough with relation to the use of ASR. Two participants claimed that they never copied errors that the support system committed, while all participants who used the support system copied one particular error.

More research into over-reliance is thus needed, especially if training modules for interpreting with ASR are to be developed.

6. Conclusions

We sought to explore three aspects of the provision of ASR support for simultaneous interpreting: (1) the viability of the support offered by InterpretBank (i.e., that numbers are displayed in an easily recognizable format, with high precision and fast enough for most interpreters' EVS; (2) the participants' interactions with the ASR support; and (3) the effects of ASR support on the participants' performance. The study, which involved six interpreting students, is limited in scope. We sought to mimic a real training environment as far as possible, and delivered speeches live in an interpreting lab with a fairly homogeneous group of participants. Given the scope of the study, the results need to be interpreted with caution and we certainly call for more studies involving larger populations of participants, including professional interpreters.

The findings suggest that the InterpretBank model used for the experiment (Model 1, with full transcription and highlighting of numbers) does generally meet ergonomic requirements. The participants were moderately satisfied with the system's usability, pointing out problems with the running transcript and the multi-stage display of the numbers, but they seemed to trust it enough to not abandon it altogether and admitted to using it for text items other than numbers. Some regarded it as a backup system for when all else fails. The system's precision was high (96%) and its latency low enough to offer interpreters a transcript before they reached the point at which they had to deliver the number.

The participants' interactions with ASR support were varied, and they consulted the output in just over half of the cases. Contradictory tendencies were found with regard to the participants' inclination to use the output after a first experience: Participants who performed well without support tended to seek support less often. The study also provided some evidence of the psychological benefits of ASR availability and of the over-reliance on ASR support.

The results also revealed that the provision of ASR improved performance: Specifically, the share of complete renditions increased in most cases and for almost all number types. However, when analyzed for each participant, a significant beneficial effect could only be found for two of the six cases. One interpreter performed less well when ASR support was available.

Apart from the limited number of participants and the fact that they were students whose performance cannot be extrapolated to the professional community, we need to mention two important lacunae among the study's limitations. First, we only analyzed rendition types for the numbers. As mentioned before, the units that accompany numbers are also affected by the cognitive load involved with the interpretation of numbers. Moreover, the overall performance throughout the speech should be investigated since it is possible that the participants, who were instructed to interpret text rich in numbers, focused their attention on these numbers and much less on the rest of the text. Overall performance should thus not be overlooked, especially when exploring the cognitive load induced by the consultation of an external resource, such as ASR output.

Second, the participants' experiences were only superficially explored in this paper. It is important for a study of human–technology interaction to survey human experiences and perceptions. Some of the results clearly indicate that the use of and benefits drawn from technological support depend on experience and expectations.

Funding

This research was partly carried out in the framework of the FWO Hercules Grant I003618N. None of the research was sponsored by InterpretBank, except for the offer to use a prototype version of the software for the purpose of the experiment. The second author contributed only to the system description in Section 3.1 and a series of comments on various topics in different sections of the text.

Acknowledgements

The authors wish to thank An Baeyens wholeheartedly for preparing and delivering the speeches.

References

Alessandrini, Maria Serena. 1990. "Translating Numbers in Consecutive Interpretation: An Experimental Study." *The Interpreters' Newsletter* 3: 77–80.

Braun, Susanne, and Andrea Clarici. 1996. "Inaccuracy for Numerals in Simultaneous Interpretation: Neurolinguistic and Neuropsychological Perspectives." *The Interpreters' Newsletter* 7: 85–102.

Brüsewitz, Nora. 2019. "Simultandolmetschen 4.0: Ist automatische Spracherkennung der nächste Schritt? [Simultaneous interpreting 4.0: Is automatic speech recognition the next step]" In *Proceedings of Übersetzen und Dolmetschen 4.0. – Neue Wege im digitalen Zeitalter* [Translation and Interpreting 4.0 – New Paths in the Digital Era], edited by Wolfram Bauer, and Felix Mayer. Berlin: BDÜ Fachverlag.

Christoffels, Ingrid K. 2004. Cognitive Studies in Simultaneous Interpreting. PhD diss. University of Amsterdam.

Collard, Camille. 2019. A Corpus-based Study of Simultaneous Interpreting with Special Reference to Sex. PhD diss. Ghent University.

Corpas Pastor, Gloria, and Lily May Fern. 2016. *A Survey of Interpreters' Needs and Practices Related to Language Technology*. Technical report [FFI2012-38881-MINECO/TI-DT-2016-1]. Málaga: University of Málaga.

Defrancq, Bart. 2015. "Corpus-based Research into the Presumed Effects of Short EVS." *Interpreting* 17 (1): 26–45. https://doi.org/10.1075/intp.17.1.02def

Desmet, Bart, Mieke Vandierendonck, and Bart Defrancq. 2018. "Simultaneous Interpretation of Numbers and the Impact of Technological Support." In *Interpreting and Technology*, edited by Claudio Fantinuoli, 11–24. Berlin: Language Science Press.

Fantinuoli, Claudio. 2017. "Speech Recognition in the Interpreter Workstation." In *Translating and the Computer 39: Proceedings*, edited by João Esteves-Ferreira, Juliet Macan, Ruslan Mitkov, and Olaf-Michael Stefanov, 25–34. Geneva: Editions Tradulex.

Fantinuoli, Claudio. 2018. "Interpreting and Technology: The Upcoming Technological Turn." In *Interpreting and Technology*, edited by Claudio Fantinuoli, 1–12. Berlin: Language Science Press.

Frittella, Francesca Maria. 2019. "'70.6 Billion World Citizens': Investigating the Difficulty of Interpreting Numbers." *Translation and Interpreting* 11 (1): 79–99. https://doi.org/10.12807/ti.111201.2019.a05

Gile, Daniel. 1995. *Regards sur la recherche en interprétation de conférence* [Views on Conference Interpreting Research]. Lille: Presses Universitaires de Lille.

Jones, Roderick. 2002. *Conference Interpreting Explained*. 2nd ed. Manchester: St Jerome Publishing.

Korpal, Paweł. 2016. Linguistic and Psychological Indicators of Stress in Simultaneous Interpreting. PhD diss. Adam Mickiewicz University.

Korpal, Paweł, and Katarzyna Stachowiak-Szymczak. 2020. "Combined Problem Triggers in Simultaneous Interpreting: Exploring the Effect of Delivery Rate on Processing and Rendering Numbers." *Perspectives: Studies in Translation Theory and Practice* 28 (1): 126–143. https://doi.org/10.1080/0907676X.2019.1628285

Lamberger-Felber, Heike. 2001. "Text-oriented Research into Interpreting – Examples from a Case-study." *Hermes* 26: 39–64.

Mazza, Cristina. 2001. "Numbers in Simultaneous Interpretation." *The Interpreters' Newsletter* 11: 87–104.

Mead, Peter. 2015. "Numbers." In *Routledge Encyclopedia of Interpreting Studies*, edited by Franz Pöchhacker, 286–288. Abingdon: Routledge.

Oléron, Pierre, and Hubert Nanpon. 1965/2002. "Research into Simultaneous Translation." In *The Interpreting Studies Reader*, edited by Franz Pöchhacker and Miriam Shlesinger, 42–51. London: Routledge.

Pinochi, Diletta. 2009. "Simultaneous Interpretation of Numbers: Comparing German and English to Italian. An Experimental Study." *The Interpreters' Newsletter* 14: 33–57.

Prandi, Bianca. 2018. "An Exploratory Study on CAI Tools in Simultaneous Interpreting: Theoretical Framework and Stimulus Validation." In *Interpreting and Technology*, edited by Claudio Fantinuoli, 25–54. Berlin: Language Science Press.

Setton, Robin, and Andrew Dawrant. 2016. *Conference Interpreting: A Complete Course*. Amsterdam: John Benjamins. https://doi.org/10.1075/btl.120

Timarová, Sarka. 2012. Working Memory in Simultaneous Interpreting. PhD diss. KULeuven.

Vogler, Nicolai, Craig Stewart, and Graham Neubig. 2019. "Lost in Interpretation: Predicting Untranslated Terminology in Simultaneous Interpretation." *ArXiv:1904.00930 [Cs]*. http://arxiv.org/abs/1904.00930

Address for correspondence

Bart Defrancq
Translation, Interpreting and Communication
Ghent University
Groot-Brittanniëlaan 45
B 9000 Ghent
Belgium

bart.defrancq@ugent.be
https://orcid.org/0000-0003-0296-0438

Co-author information

Claudio Fantinuoli
Mainz University

fantinuolini@mainz.de
https://orcid.org/0000-0003-1312-0741

Publication history

Date received: 22 September 2019
Date accepted: 28 October 2020

Author Query

- Please provide a complete author address '(Claudio Fantinuoli)' in this article.