The impact of visible lip movements on silent pauses in simultaneous interpreting

Anne Catherine Gieshoff
Zurich University of Applied Sciences

Abstract

Simultaneous interpreting requires interpreters to listen to a source text while producing the target text in a second language. In addition, the interpreter needs to process various types of visual input, which may further increase the already high cognitive load. A study with 14 students of interpreting was conducted to investigate the impact of a speaker's visible lip movements on cognitive load in simultaneous interpreting by analysing the duration of silent pauses in the target texts. Background noise masking the source speech was introduced as a control condition for cognitive load. Silent pause durations were shorter when interpreters saw the speaker's lip movements, which indicates that interpreters benefitted from visual input. Furthermore, silent pause durations were longer with noise, which suggests that comparative silent pause durations can indicate changes in cognitive load.

Keywords: simultaneous interpreting, visual input, cognitive load, silent pauses, lip movements

1 Introduction

Conference interpreters need to process multiple sources of sensory input simultaneously. Apart from auditory information such as the source speech and their own production, they also receive a great deal of visual information. Interestingly, interpreters do not seem to experience visual input as an additional burden, but instead regard visual input and, especially a view of the speaker, as valuable and helpful, if not indispensable, for successfully rendering the source speech into the target language (Bühler 1985). Visual input, however, can take on different forms that may have different effects on the interpreters' cognitive load. To date, a systematic investigation of the impact of visual input on simultaneous interpreting (SI) is lacking. One type of visual input that could facilitate SI by enhancing speech perception is the speaker's lip movements. In the study presented in this article, I tested whether lip movements have a positive effect on interpreters' cognitive load by measuring the duration of silent pauses in the interpreters' renditions. As theorized by some authors, long silent pauses could reflect an attentional shift either towards source-speech analysis or towards speech production and might therefore indicate increased cognitive load. Background noise overlaid on the source speech was introduced as a control variable to ensure the validity of the method.

2 Theoretical background

The following sections provide an overview of the literature that sparked the study. The first section reviews studies on visual input in SI and remote interpreting and develops the hypothesis on the effect of the speaker's lip movements in SI. The second section deals with the effect of background noise in SI. The section concludes with a section on disfluencies in SI and explains why silent pauses in particular may be interesting with regard to cognitive load.

2.1 Visual input in simultaneous interpreting

Simultaneous interpreting is a capacity-consuming task and it is often assumed that interpreters work close to the saturation point of their attentional resources (Gile 2009; but see Seeber 2015 for a different account). One aspect that makes SI so demanding is the multitude of sensory information that needs to be processed at the same time in order to render the source speech into the target language. Sitting in her booth, the interpreter takes in not only auditory information, but also visual information: the lip movements and gestures of the speaker, presentation slides, glossaries or manuscripts and handwritten notes, to name but a few (for typical visual input during SI see, for instance, Seubert 2017). The question that arises is whether this multitude of visual input increases the already heavy cognitive burden of the interpreter or – in contrast – facilitates the task, for example, by providing complementary information that may help to retrieve lost segments.

Research that may shed some light on the impact of visual input includes studies in which visual input has been completely blocked or limited, as in remote interpreting. The picture that emerges from these studies is inconsistent: on the one hand, interpreters reported increased fatigue and concentration difficulties when their view of the speaker was completely blocked (Rennert 2008) or limited, as in remote interpreting situations (Moser-Mercer 2003, 2005; Roziner & Shlesinger 2010). In line with interpreters' reports, Moser-Mercer found interpreting performance to decline faster in one study on remote interpreting (Moser-Mercer 2003, 2005). On the other hand, the majority of studies conducted to date seem to suggest that overall interpreting performance does not suffer from the absence of visual input, neither in remote interpreting with limited visual input (Roziner & Shlesinger 2010) nor in an experimental condition where the view was completely blocked (Anderson 1994; Rennert 2008).

Interestingly, visual input has been dealt with only marginally in models of SI. Most processing models of SI do not mention visual input explicitly (see, for instance, Gerver 1975; Mizuno 2005; Moser 1978). One exception is Setton (1999), who includes both the environment and the speaker's lip movements and gestures as part of the source-text analysis. It is, however, unclear, where other types of visual input play a role. Models focusing more on the communicative aspect of SI (for instance, Pöchhacker 2005; Poyatos 1984; Rackow 2013) acknowledge the importance of visual input, but they do not discuss the cognitive load involved in processing visual items.

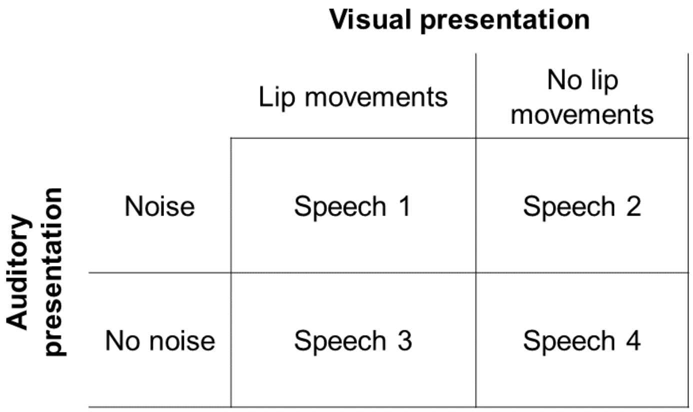
The only model that deals explicitly with visual input and cognitive load is Seeber's (2011, 2017) cognitive load model. Seeber suggests that the cognitive load experienced during SI depends very much on the type of visual input and to what extent the processing of the visual input interacts with other ongoing subtasks such as auditory processing of the speech or producing the target speech. Visual input that duplicates the auditory input and requires additional resources, such as written manuscripts, will increase the cognitive load, whereas visual input that is complementary to the auditory input, such as the speaker's lip movements or gestures, will reduce the cognitive burden because it interferes far less with the subtasks necessary to processing the source speech (Seeber 2017). The multitude of visual inputs an interpreter receives during the task may indeed be one of the reasons why previous studies on visual input in SI failed to find a consistent effect on interpreting performance: while some types of visual input may have increased the interpreter's cognitive burden, others may have lowered it or at least maintained it at a stable level. In order to avoid similar confounding effects, I decided to limit visual input in my study to the speaker's lip movements.

I expected that seeing the speaker's lip movements concurrently with the auditory input should have a facilitatory effect on SI. Moreover, I hypothesized that the contribution of visible lip movements to speech perception should be even stronger when the auditory signal is degraded. These hypotheses were motivated not only by the predictions of Seeber's (2017) cognitive load model, but also by psycholinguistic studies. Researchers observed that speech perception benefits from seeing the speaker's (congruent) lip movements (Bernstein et al. 2004; Brancazio et al. 2006; Thomas & Jordan 2004; von Kriegstein et al. 2008), and in particular in adverse listening conditions, as is the case when listening to speech under cognitive load (Mattys & Wiget 2011) or with noise (Bernstein et al. 2004; Macleod & Summerfield 1987; Vatikiotis-Bateson et al. 1998) and for hearing-impaired participants (Kramer et al. 1997). The more the auditory signal is affected, the more the visual signal contributes to speech perception (Macleod & Summerfield, 1987; Vatikiotis-Bateson et al. 1998).

To account for these observations, Massaro and Cohen (1999) developed the *Fuzzy Logical Model of Perception*. They assume that neither auditory speech nor visual speech inputs are unambiguous and that the combination of both signals helps to determine the phoneme by reducing the noise on both signals (Massaro & Cohen 1999). The phonemes /m/ and /n/, for example, may sound similar, but they are easily distinguished by the visual features (closed vs open lips). The visual signal can thus be said to be complementary to the auditory signal in the sense that the visual signal helps to 'fill in' the gaps in the auditory signal.

2.2 Manipulating speech perception with background noise

Many psycholinguists have investigated the contribution of the visual signal to speech perception by overlaying background noise on the auditory signal in order to manipulate the signal-to-noise ratio (SNR). The underlying rationale is that background noise partially covers the auditory stream and makes it less intelligible and that participants need to rely more on visual cues (Mattys et al. 2009). Background noise is not without practical relevance to conference interpreters: as suggested by McAllister (2000), noise may have disruptive effects on speech comprehension that are similar to those of accented speech to which conference interpreters seem to be exposed more and more frequently (Albl-Mikasa 2010) (for the effects of foreign accent on SI, see I-hsin et al. 2013; Sabatini 2000). Moreover, background noise has already been found to affect interpreting performance. Gerver (1974) tested the effect of noise during SI and found more errors and omissions when the SNR of the source speech was low. For this reason, in the present study, noise was introduced as a second variable. The experimental design therefore included four conditions: interpreting with no visible lip movements or noise, interpreting with visible lip movements but without noise, interpreting without visible lip movements but with noise and interpreting with visible lip movements and noise (see Figure 1).



Visual presentation

		Lip movements	No lip movements
Auditory presentation	Noise	Speech 1	Speech 2
	No noise	Speech 3	Speech 4

Figure 1. Experimental conditions.

As mentioned previously, confounding factors from the multitude of visual inputs may be one of the reasons why studies on visual input in SI have not found any clear-cut effect despite interpreters' claiming how helpful visual input is. Another reason for the mismatch between interpreters' reported experience and researchers' observations could be of a methodological nature: it is possible that the methods used so far were not sufficiently sensitive to capture subtle changes in cognitive load. In previous studies, researchers concentrated on interpreting performance and assessed whole renditions or excerpts by asking judges either to score the renditions – based either on a scale (Anderson 1994; Roziner & Shlesinger 2010) or on the numbers of errors and omissions (Moser-Mercer 2005) – or to analyse them qualitatively (Rennert 2008). Most often, researchers covered essential aspects of interpreting performance such as informativeness and accuracy (Anderson 1994) or errors and word choice[1] (Roziner & Shlesinger 2010). Moser-Mercer (2005) is an exception: she derived an overall error score by analysing renditions according to several aspects that included meaning errors as well as less

[1] The authors mention only two of six dimensions that were used for ratings.

substantial parameters such as grammar, style and prosody. Meaning errors received the highest weighting of all the parameters. Still, she was the only one to find an effect on interpreting performance in remote interpreting as opposed to on-site interpreting.[2]

Serious errors and omissions, however, may not always fully reflect subtle changes in cognitive load because interpreters regard them as a crucial criterion for interpreting quality (Bühler 1986; Zwischenberger 2010) and may adapt their efforts to avoid them. In other words, when conditions become more challenging – for instance, because visual input is lacking, and the interpreter feels unable to satisfy all aspects of interpreting quality to the same extent as usual – s/he may decide to concentrate on sense consistency and completeness at the cost of less substantial aspects. It is then only at this more fine-grained level that differences in cognitive load can be observed. One of these 'less substantial' aspects could be disfluencies. Fluency is regarded as an important, but – crucially – not *the* most important criterion for assessing interpreting quality when judged by listeners (Pradas Macías 2006; Rennert 2019; Yu & van Heuven 2017) and interpreters alike (Bühler 1986; Chiaro & Nocella 2004; Zwischenberger 2010). Disfluencies might therefore reflect changes in cognitive load more readily than serious errors and omissions.

2.3 Disfluencies as an indicator of cognitive load

Disfluencies include a wide variety of phenomena in interpreted speech. One of the most basic distinctions is filled pauses and silent (unfilled) pauses. Filled pauses that are found most frequently in interpreting studies include hesitations (Cecot 2001; Lin et al. 2018; Plevoets & Defrancq 2016, 2018; Rennert 2019; Tissi 2000; Yu & van Heuven 2017), vowel lengthening (Cecot 2001; Lin et al. 2018; Rennert 2019; Tissi 2000; Yu & van Heuven 2017), and repetitions of words, corrections and false starts (Cecot 2001; Tissi 2000; Yu & van Heuven 2017). Silent pauses can be described as interruptions in renditions without hesitations, corrections, false starts or other sounds such as coughing (Cecot 2001; Rennert 2019; Tissi 2000).

Much research on disfluencies in interpreting focuses on disfluency patterns (Ahrens 2004; Chmiel et al. 2017; Goldman-Eisler 2002; Tissi 2000) and/or investigates how disfluencies in the target text affect interpreting quality (Pradas Macías 2006; Rennert 2019; Yu & van Heuven 2017) or how disfluencies in the source text influence the interpreter (Cecot 2001). But

[2] The author does not report the detailed results for each parameter. It is therefore not clear how each parameter influenced the total error score.

disfluencies have also been used to investigate cognitive load during SI. In a recent corpus study, Plevoets and Defrancq (2018) found that high lexical density and formulaic expressions significantly affected the number of filled pauses in interpreted and non-interpreted speech and concluded that those two factors can cause changes in cognitive load.

As the examples above show, researchers typically analyse the frequency of disfluencies, that is, how often certain types of disfluency occur under a given condition. The case is a little different for silent pauses, where either duration (Gerver 2002) or both frequency and duration of silent pauses are investigated (Cecot 2001; Tissi 2000; Yu & van Heuven 2017). In addition, in some cases the total time of silent pauses (Tissi 2000) or the pause ratio (Chmiel et al. 2017) is indicated. Typically, researchers find fewer silent pauses in interpreted compared to natural (Tissi 2000) or shadowed speech (Chmiel et al. 2017) or more generally in conditions that are associated with a higher load on speech comprehension (Cecot 2001; Chmiel et al. 2017; Lin et al. 2018). At the same time, the duration of silent pauses in interpreted speech (Tissi 2000) or under high load is longer (Gerver 2002).

Silent pauses may not be related to higher load only. Early work by Goldman-Eisler (1968) suggested that very short pauses (<250 ms) were of an articulatory nature. Moreover, silent pauses also play an important role in marking the end of information units (Ahrens 2005; Cecot 2001; Goldman-Eisler 2002). Ahrens (2004) thoroughly analysed interpretations by three interpreters from the same source text and found a stronger segmentation in the interpretations than in the source speech. It seemed that especially short silent pauses (<0.4 seconds) are used to mark the boundary of an information unit (Ahrens 2004: 151–163). These examples suggest that short silent pauses may be misleading when a researcher investigates cognitive load since they are most probably linked to physiological or communicative processes.

Different authors have theorized about the cognitive processes that underlie disfluency phenomena, and in particular filled and silent pauses. One of the first hypotheses comes from Goldman-Eisler (1958, 1961), who suggested that long silent pauses in spontaneous speech are caused by verbal planning. Based on a corpus of Chinese–English interpreting, Setton (1999) presumed that long filled pauses occur with attentional shifts towards target text production whereas long silent pauses indicate exclusive attention to the source speech. According to the author, short pauses, both silent and filled, seem to reflect a more balanced attention between source text analysis and target text production (Setton 1999: 245–248). Ahrens shared the view that long silent pauses could reflect source text analysis (Ahrens 2004: 163–172). In a related publication, however, the author points out that long silent pauses may also occur when interpreters wait for new input (Ahrens 2005: 163–172). The interpreter may use waiting as a strategy to gather more information that helps produce the next information unit in the target language. This may therefore in some cases point to difficulties in source speech processing. But long silent pauses may also occur on other occasions – for instance, when the speaker is searching for a word.

It is against this tentative background that I was motivated to test whether silent pauses could indicate differences in cognitive load during SI while the speaker's lip movements are either visible or not. As mentioned above, lip movements are part of the visual input that interpreters receive and may facilitate speech comprehension. The absence of visible lip movements, in turn, may impede source text comprehension and in consequence also affect target speech production. Hence, I expected silent pauses to be shorter with visible lip movements than without. Regarding the second manipulation – noise overlaid on the source speech – I expected longer silent pauses in the condition with noise than in the condition without noise. This is because noise would obscure different parts of the source speech. As a result, the interpreter would need to wait longer to receive sufficient input before starting their interpretation. Finally, as demonstrated by Vatikiotis-Bateson et al. (1998), and Macleod and Summerfield (1987), I hypothesized that these two factors – visible lip movements and noise – would interact in a way that interpreters would benefit more from visible lip movements during SI with noise than without noise. Consequently, the difference in silent pause duration during SI with and without noise would be smaller when lip movements are visible than without lip movements.

3 Empirical study

The study presented here was part of a larger research project. This article is limited to the results of the analysis of silent pause durations and subjective reports. The results of further analyses are reported in Gieshoff (2018).

3.1 Participants

In total, 14 students of conference interpreting in their final year gave informed consent for their participation in the study after receiving information about the procedure of the experiment and the data that were to be collected. All the participants had received at least three semesters of training in SI and had German as their native language. English was the B or C language for all the participants. The number of participants was limited for practical reasons. One participant was completely excluded from the analysis because the recording quality was insufficient for silent pause extraction. A total of 52 recordings, four from each participant, were collected. All the participants confirmed that they felt well and in good health and received €10 in exchange for their participation.

3.2 Material

The experiment consisted of two parts: a pre-test and the actual main experiment. The material for the pre-test comprised 64 English nouns selected from among the 5,000 most frequently used words of American English (Davies 2008). All words were concrete and monosyllabic. They were spoken by an American native-speaker and recorded as sound files. All the words were grouped into four lists of 16 words each. Each list was then overlaid with noise at four different signal-to-noise ratios, ranging from level 0.1 (10% of the maximal sound level of the sound card) to level 0.4 (i.e., 10% to 40% of the maximal sound level).[3] The SNR that was applied to each list was randomized between participants.

The material for the main experiment consisted of four speeches of approximately 590 words (*M* = 588, *SD* = 5.23) covering four different topics: air travel, the Greek economic crisis, working conditions and demographic change. They were initially chosen from the basic level corpus of the speech repository of the European Union (European Commission 2009a, 2009b, 2012a, 2012b), but large parts were rewritten, edited or deleted in order to obtain four speeches that are as comparable as possible. The structure of the first paragraph[4] was the same across all speeches and served as a 'warm-up' for the interpreters. It contained the usual introductory

[3] For the aforementioned reasons related to sound objects in Psychopy (Peirce 2007), the signal-to-noise ratio is not expressed in decibels but was based on the volume scale in Psychopy.

[4] The structure of the first paragraph was as follows: 'Ladies and Gentlemen, I am very pleased to be here at the international conference for [topic]. I am honored to speak to so many distinguished guests who [short description with reference to the topic]. They are an example for all of us. Today, I want to talk about [topic].'

expressions and announced the topic of the text twice. Only the 5,000 most frequently used words of American English (Davies 2008) were used. The speeches did not contain any potential local problem triggers such as numbers or proper names. All of the sentences were written in the active mode and had no more than one subordinate clause so that all sentences contained on average 12.5 words (*SD* = 2.2). The number of function words (articles, prepositions and other words with a purely grammatical function) and the type–token ratio served as indicators of information density. In every text, function words made up approximately 40% of all words (function-word ratio: *M* = 0.4, *SD* = 0.03; type–token ratio: *M* = 0.48, *SD* = 0.05). Beyond this quantitative evaluation of information density, the speeches were edited in order to reduce information density at the textual level and to allow the interpreters to catch up in case they missed the message. For instance, essential messages were repeated in different words. Filler sentences containing evident information without impact on text coherence were introduced. Finally, all the underlying logical relationships within the text were made explicit through the use of conjunctions.

The speeches were read out by the same male American native-speaker and recorded on video. Each of the videos was about four minutes long. The speech rate was kept constant at a rate of 140 words per minute within and between texts. The mean duration of silent pauses in the source text was 845 milliseconds (*M* = 0.845 s, range: 0.505–1.863 s). The video showed the speaker's whole face as a video stream so that the participants would be able to see the speaker's lip movements while he was giving his speech. The lip movements were congruent with the auditory stream. In order to ensure that the video would be replayed smoothly without interruptions, the resolution was down-sampled to a sampling rate of 16,384 kbits/s with a resolution of 1,920 × 1,080 and 29.97 frames per second. Sound was played in stereo at a volume level of 0.1[5] with a sampling rate of 320 kbits/s. In the condition without lip movements, the sound stream remained the same, but the video stream was replaced by a freeze frame of the speaker's face in the same resolution as the video (1,920 × 1,080). In the noise condition, white noise, that is, noise with different frequencies at the same intensity, was added to the audio stream, while in the no-noise condition, speeches were presented without any interfering noise. In

[5] In Psychopy (Peirce 2007), the volume of a sound object can be set on a scale from 0.0 (silent) to 1.0, where 1.0 is the maximum volume of the sound card. For this reason, the volume is not given in dB but based on Psychopy's volume scale.

order to reduce potential speech-related effects, I created two counter-balancing groups (groups A and B) and switched the speeches in the condition with and without visible lip movements (see Figure 2). The order of presentation was randomized.

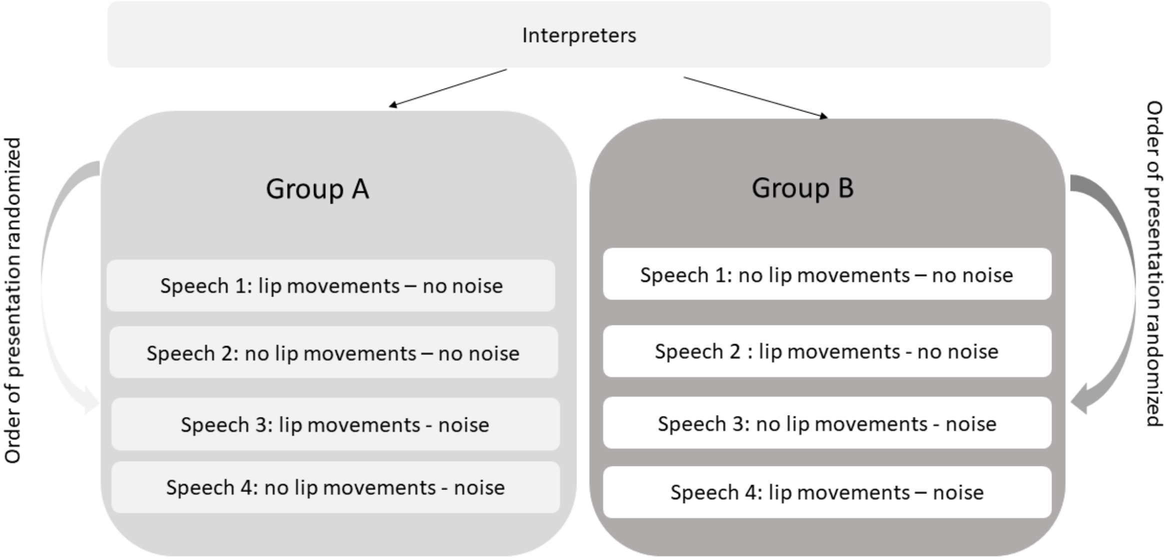

Interpreters
Group A
Speech 1: lip movements – no noise
Speech 2: no lip movements – no noise
Speech 3: lip movements - noise
Speech 4: no lip movements - noise
Group B
Speech 1: no lip movements – no noise
Speech 2 : lip movements - no noise
Speech 3: no lip movements - noise
Speech 4: lip movements – noise
Order of presentation randomized
Order of presentation randomized

Figure 2. Counterbalanced experimental design

3.3 Procedure

The experiment consisted of two parts: a pre-test and the main experiment. The entire procedure took place in the laboratory of the Faculty of Translation Studies, Linguistics and Cultural Studies of the University of Mainz and lasted about 40 minutes; the main experiment lasted about 30 minutes. All the instructions were given in German, the participants' native language. The purpose of the pre-test was to adjust the SNR for each participant to their individual 75% word recognition threshold. In order to do so, the participants were presented aurally with 64 random but concrete and frequently used English words (see material section). After each word had been presented, the participants typed the word they had recognized. They could correct their answers before submission, but it was not possible to hear the stimulus again. After the pre-test, an algorithm computed the number and percentage of correct guesses for each SNR and subsequently selected an individual 75% threshold for each participant. This threshold was defined as the lowest signal-to noise ratio with a correct recognition rate of at least 75%. As 16 words were presented for each SNR, this percentage corresponded to at least 12 correctly identified words. The resulting SNR from the pre-test was then applied to the source speeches in the condition with noise. The mean ratio of correct guesses for each SNR is given in Table 1. As the participants reached only the two highest SNRs, noise level 0.1 and noise level 0.2, the former will be referred to as 'high SNR' and the latter as 'low SNR'.

Table 1. Mean ratio of correct word recognition for each signal-to-noise ratio and its standard deviation

Signal-to-noise ratio	Mean ratio of correct word recognition in %	Standard deviation	Number of participants
0.1	80.4	8.79	8
0.2	67.5	10.62	5
0.3	49.2	12.91	0
0.4	32.9	11.92	0

Note: The right column indicates the number of participants to whom the respective SNR was applied.

After the pre-test, the main experiment with four trials started. Each trial included one of the speeches so that every participant interpreted all of the speeches. The participants started the source speech by pressing a key. After the interpreting task, they were asked to rate the video and audio quality, the speech complexity and the speech rate on a four-point scale. Video and audio quality ratings[6] were elicited to ensure that the experimental conditions – that is, the noise vs the no-noise condition and the condition with vs without lip movements – were clearly distinct. Text difficulty[7] and speech-rate ratings[8] served to ensure that the speeches were perceived as being comparable in terms of speech rate and difficulty. After the whole procedure, the participants

[6] The exact wording of the question was: 'Wie empfanden Sie die Bildqualität?' ['How did you find the image quality?'; my translation]. For the sound-quality rating, the word *Bildqualität* was replaced by *Tonqualität* [sound quality]. Options were as follows: 'ausgezeichnet', 'gut', 'mittelmässig', 'schlecht' ['excellent', 'good', 'average', 'bad'; my translation].

[7] The exact wording was: 'Wie schwierig fanden Sie den Textinhalt?' ['How difficult did you find the speech content?'; my translation]. The options were 'sehr leicht', 'leicht', 'mittelmässig', 'schwer' ['very easy', 'easy', 'average', 'difficult'; my translation].

[8] The exact wording was: 'Wie empfanden Sie die Vortragsgeschwindigkeit?' ['How did you find the speech rate?'; my translation]. The options were 'langsam', 'angemessen', 'flott, aber machbar', 'zu schnell' ['slow', 'appropriate', 'quick, but doable', 'too fast'; my translation].

were invited to report orally any technical problem or difficulty they had experienced. No difficulties were reported.

3.4 Data analysis

3.4.1 Subjective reports

The participants' ratings were transformed into numerical values ranging from 1 ('bad'/'difficult'/'too fast') to 4 ('very good'/'very easy'/'slow'). I conducted a paired Wilcoxon signed rank test on the video- and sound-quality ratings to test whether the levels of the two predictors, visual and auditory presentation, were sufficiently distinct. The participants' individual SNR depended on their performance in the pre-test and therefore differed between participants (high or low SNR). In order to be sure that even for the high SNR the noise and no-noise conditions were still sufficiently distinct, I conducted separate Wilcoxon signed rank tests for both of the signal-to-noise ratios. Regarding the speech difficulty and speech rate ratings, I conducted a Friedman signed rank test in order to check whether there were significant differences between the four speeches. All the analyses were carried out in R version 3.3.2 (R Core Team 2016) with the package *coin* (Hothorn et al. 2008).

3.4.2 Silent pause durations

Silent pauses from each recording were automatically extracted using *Praat* (Boersma & Weenink 2013). As short pauses to mark word or sentence boundaries are normal in speech, only silences longer than 500 milliseconds were considered. The reasons for this choice were twofold: first, research by Ahrens (2004: 151–163) suggests that pauses of less than 0.4 seconds are used to mark the end of an information unit. Second, an analysis of the source speech showed that the shortest silent pause in the source speech was 0.5 seconds long (*M* = 0.845 s, range: 0.505–1.863 s). Further, all observations five seconds after the beginning and five seconds before the end of the recording were removed as the interpreter needs to wait until the first few segments before starting the interpretation of the speech. After the removal of four observations that appeared to be invalid, the overall mean duration of silent pauses was 1.481 seconds (*MD* = 1.057 s, range: 0.500–11.880 s). The mean duration for the experimental condition with noise was slightly higher (*M* = 1.638, *MD* = 1.141) than without noise (*M* = 1.346, *MD* = 0.997) and also slightly higher for the condition without lip movements (*M* = 1.577, *MD* = 1.135) compared to the condition with lip movements (*M* = 1.394, *MD* = 0.973).

In order to test whether these differences were statistically significant, I constructed a linear mixed effects model using the *lme4*-package (Bates et al. 2015) in R version 3.3.2 (R Core Team 2016). The choice to conduct a mixed model was made in order to account for idiosyncratic variations that were either related to the participant or to the source speech. As underlying distribution, I assumed a normal distribution. However, given the structural similarity between reaction time data and silence pause duration, I followed the recommendation by Lo and Andrews (2015) and, in addition, reconstructed the same model with an exponential distribution as a generalized mixed effects model. The dependent variable is the duration of silent pauses in the participants' renditions of the source speech. Random effects included intercepts for participants and speeches. Fixed effects included the main effects of visual presentation, auditory presentation and SNR, as well as an interaction of SNR and auditory presentation, and visual and auditory presentation. Visual and auditory presentation as well as SNR were categorical variables with two levels each: lip movements vs no lip movements for the predictor visual presentation; noise vs no noise for the predictor auditory presentation; and high vs low for the predictor SNR. *P*-values for fixed effects were estimated indirectly by comparing the full model containing all the predictors with likelihood ratio tests against a reduced model without the fixed effect in question (Bates et al. 2015: 35). When the main effects were tested, the corresponding interaction terms were equally removed from the full model because the interaction without the main predictor is not meaningful. Confidence intervals for plots were obtained with the R-package *effects* (Fox & Weisberg 2018). Plotting was performed with the R-package *ggplot2* (Wickham 2009).

3.5 Results

3.5.1 Subjective reports

According to a Wilcoxon signed rank test, video-quality ratings [*MD* (lip movements) = 2, *MD* (no lip movements) = 1] differed significantly between the conditions with and without lip movements ($Z = 4.133$, $p < 0.001$). Similar results were found for the conditions with and without noise: the audio-quality ratings differed significantly between the conditions with and without noise [*MD* (no noise) = 2, *MD* (noise) = 1]. This was the case both for the low SNR (*Z* = 2.763, *p* <0.01) and for the high SNR (*Z* = 3.127, *p* <0.01). A Friedman test indicated no differences in the distribution of the speech-rate ratings (*MD* = 3, $\chi^2(3)$ = 5.927, *p* >0.1) or the speech-difficulty ratings (*MD* = 3, $\chi^2(3)$ = 1.779, *p* >0.1) between the four speeches.

3.5.2 Silence durations

Estimates, standard error and *z*-value of the final model, in addition to the statistics of model comparisons for each predictor, are reported in **Fehler! Verweisquelle konnte nicht gefunden werden.**. Main effects were found for visual presentation (*F(9,2)* = 26.488, *p* <0.001), for auditory presentation (*F(9,3)* = 77.466, *p* <0.001) and for SNR (*F(9,2)* = 73.704, *p* <0.001). The negative estimate of the predictor visual presentation (reference level: no lip movements, *Estimate* = −0.209, *SE* = 0.040*, z(3154)* = −5.176) suggests that the visibility of lip movements contributed to decreasing the duration of silent pauses. The positive estimates of the predictor auditory presentation (*Estimate* = 0.058, *SE* = 0.058*, z(3154)* = 1.004) and signal-to-noise ratio (*Estimate* = 0.117, *SE* = 0.166*, z(3154)* = 0.706) suggests that silent pause duration increased in the condition with noise and with lower signal-to-noise ratios. The rather low estimates, however, seem to suggest that their effect on silent pause duration is marginal. The estimate for interaction between auditory presentation and SNR was again positive, suggesting that the effect of noise is stronger for a lower SNR (*Estimate* = 0.686, *SE* = 0.083*, z(3154)* = 8.236). The interaction of auditory and visual presentation failed to reach significance. **Fehler! Verweisquelle konnte nicht gefunden werden.** and 4 depict the effects of the final model. The same model with an exponential distribution yielded similar results for visual presentation (visual presentation: *Estimate* = −0.186, *SE* = 0.030, *t(3154)* = −6.187, *p* <0.001; *Estimate* = 0.0535, *SE* = 0.039, *t(3154)*=1.383, *F(8,2)* = 96.925; SNR: *Estimate* = 0.128, *SE* = 0.154, *t(3154)* = 0.30, *F(8,2)* = 95.65, *p* <0.001; interaction SNR and auditory presentation: *Estimate* = 667, *SE* = 0.072, *t(3154)* = 9.280, *F(8,1)* = 90.376, *p* <0.001).

Table 2. Estimates, standard error and *t*-value for predictors of the final model and Chi-square, degrees of freedom and *p*-value for each model comparison

	Final model			*Model comparison*		
Predictor	*Estimate (β)*	*SE*	*z(3154)*	*F*	*DF*	*p* χ^2

Visual presentation	−0.209	0.040	−5.176	26.488	9,2	<0.001
Auditory presentation	0.058	0.058	1.004	77.466	9,2	<0.001
Signal-to-noise ratio	0.117	0.166	0.706	73.704	9,3	<0.001
Auditory presentation * signal-to-noise ratio	0.686	0.083	8.236	67.093	9,1	<0.001
Auditory * visual presentation ratio	Not significant and therefore not included in the final model			0.022	9,1	0.882

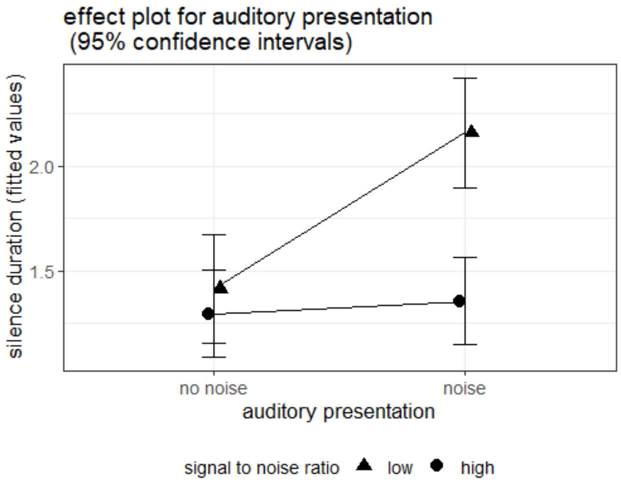


Figure 3. Effect of auditory presentation. The plot indicates the SNR: triangle = low SNR, circle = high SNR.

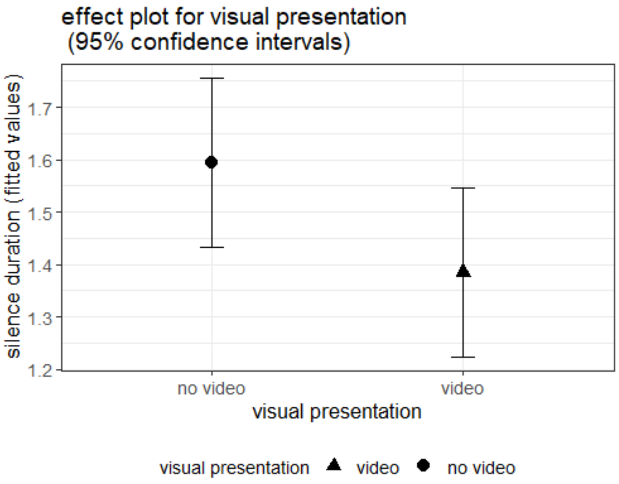


Figure 4. Effect of visual presentation. The plot indicates respectively the condition without lip movements (circle) and with lip movements (triangle).

4 Discussion

Their subjective ratings suggested that the participants perceived the speeches as being comparable in terms of speech rate and speech complexity. It also appeared that the conditions (lip movements vs no lip movements, noise vs no noise) were sufficiently distinct to be noticed by the participants even though the sound and image quality were not optimal. The reasons for the participants' low ratings of the video quality are probably of a technical nature: the videos needed to be down-sampled in order to avoid interruptions when they were played during the experiment.

The hypothesis stating that silence durations during SI with visible lip movements should be shorter than without visible lip movements was corroborated. This is in line with the *fuzzy logical model of speech perception* (Massaro & Cohen 1999) and also with Seeber's cognitive load model (Seeber 2011, 2017) or other models of SI that assume an overall lower cognitive load when demands in speech comprehension decrease (see Chernov 1994; Gerver 1975; Gile 2009). It also extends findings from psycholinguistic studies suggesting that visible lip movements facilitate speech perception (Benoit et al. 1994; Thomas & Jordan 2004; von Kriegstein et al. 2008) to simultaneous interpreting. The study also revealed that noise can lead to longer silent pause durations if the signal to noise ratio is sufficiently low. These results enhance the findings of Gerver (1974) in that they suggest that noise at a low SNR leads not only to a higher number of errors and omissions, but also to longer silent pause durations. It may be speculated that there is a causal relationship between both findings: when noise masks the source speech and listening comprehension is hampered, the interpreter needs to wait for more input to 'fill the gaps'. As a result of the higher memory load the interpreter might more easily omit some information or render it incorrectly. As regards the higher SNR, it did apparently not disrupt listening comprehension sufficiently to affect the silent pause durations in the interpretations.

The interaction of visual and auditory presentation was not significant. The interpreters did not benefit more from seeing the speaker's lip movements in the noise condition than in the condition without noise. The reasons for this may be twofold. One possible explanation is that the benefit of seeing the lip movements is limited. This means that lip movements may not compensate fully for degraded auditory input but only to a certain degree. This explanation is supported by an eye-tracking study by Vatikiotis-Bateson et al. (1998) which demonstrated that even at the lowest signal-to-noise ratios, participants’ fixations on speakers’ lip movements made up only about 50% of all fixations. Another possible reason why no interaction has been observed is related to the experimental design: based on the results of the pre-test, only the two highest signal-to-noise ratios were applied (noise level 0.1 = ‘high’ and noise level 0.2 = ‘low’). Even though the lower SNR affected the silent pause duration, it may not have been disruptive enough to impede listening comprehension in a way that (partial) lip reading became necessary for the participants. It would certainly be interesting to conduct a study with lower signal-to-noise ratios to see whether the benefit of lip movements increases when the speech is more severely degraded.

4.1 Limitations of the study

Several points should be kept in mind for this study. First, it is a study of a more explorative nature and the number of participants is fairly low. It might be interesting to see whether these findings can be replicated with a higher number of participants and extended to other factors that are potentially problematic in SI and which affect speech comprehension, such as accent. Moreover, the study presented here used students of interpreting as participants. It is not clear whether the same pattern for silent pauses would be observed in professional conference interpreters. First, it seems safe to assume that professional interpreters have better linguistic skills, more experience with similar topics or simply more world knowledge that enables them to anticipate incoming information more easily and therefore to avoid silent pauses. Second, fluency in the target text has been found to be an important vector of listeners’ acceptance (Pradas Macías 2006; Yu & van Heuven 2017). Professional interpreters may be more aware of the importance of fluency in the target text than students of interpreting, and may make use of different strategies to ensure a fluent target text. Finally, the analysis focused on silent pause duration. Although it seems plausible to assume that long pauses of several seconds indicate a disruption of the interpreting process and an attentional focus on the source speech, it should be noted that this analysis contains no information about the linguistic function or cognitive process that would be responsible for triggering a silent pause. Such an investigation would at least require knowledge of the context – for example, the position of the pause in the target sentence and the source sentence.

4.2 Potential and limitations of silent pauses in the target text as cognitive load indicator

Even though research suggests that silent pauses are perceived as negative by the listener (Pradas Macías 2006; Rennert 2019) and may even bias accuracy judgments in consecutive interpreting (Yu & van Heuven, 2017), it is important to note that this study is not intended to be a contribution to the discussion on interpreting quality. This would have required asking listeners about their perception of the target text. Instead, one purpose of the present study was to investigate how well silent pauses can indicate an increased demand or cognitive load for visual input – here limited to the speaker's lip movements – in SI. Background noise, a factor that has already been shown to increase the number of errors in SI (Gerver 1974), served as a control.

The choice to investigate silent pause durations rather than other types of disfluency such as filled pauses was made mainly for practical reasons. The most important reason is that silent pauses can be retrieved automatically without (time-consuming) manual tagging. This is different from other types of disfluency, which are usually counted or tagged on the sound file to determine their frequency. At the beginning of the study, however, it was not certain whether and how well silent pause durations could indicate cognitive load. For this reason, I introduced noise as a control variable. I observed a similar effect of noise on silent pause durations as Gerver (1974) did on errors and omissions: both increased with decreasing SNR. This observation speaks in favour of using comparative silent pause duration as an indicator of cognitive load and may be a useful finding for further studies on cognitive load in conference interpreting.

However, this method has also some limitations and sources of error that I would like to highlight. One source of error certainly pertains to silent pauses that are of a physiological or a communicative nature. Excluding very short silent pauses and taking into account the position of silent pauses within the target text may help to obtain more reliable and accurate results. Another point to keep in mind is source-speech characteristics and the pace at which a speech is delivered. A high number of interruptions and disfluencies in the source speech, which is often the case in authentic speeches, can in themselves increase cognitive load and affect silent pause durations in the interpretation. This effect can interact with the actual variable being studied and can complicate the interpretation of the results as the observed effects are not clearly attributable to the experimental manipulation anymore. The analysis of silent pauses therefore seems better suited to source speech material that is recorded at a constant pace, as was the case in the present study, than to spontaneous speech and, accordingly, better suited to controlled 'laboratory' experiments than ecologically more valid real-life settings. Finally, interpreters may also choose to use strategies to avoid silent pauses and to deliver a fluent interpretation. To do so, they may, for instance, slow down their speech rate, reformulate a segment or introduce an additional neutral sentence. A closer look at the interpretation itself and any strategies that have been used may yield complementary (but certainly not exhaustive) insights into the nature of the silent pause. Furthermore, it is certainly helpful to combine disfluency analyses with subjective reports from the participants in order to make sure that they perceive an experimental task that has been designed to represent high load to be more difficult or more strenuous.

5 Conclusion

During SI, the interpreter processes not only auditory input, but also various types of visual input that may have very different effects on the interpreters' cognitive load, depending on how much they interact with the ongoing cognitive processes required for the auditory input (Seeber 2017). According to Seeber's (2017) cognitive load model for SI, visual input that is complementary to the auditory input, such as the speaker's lip movements, gestures or facial expressions, should facilitate speech comprehension. However, research in interpreting studies has not yet been able to find a clear facilitating effect for complementary visual input, possibly because of confounding effects or because the methods are not sufficiently sensitive. This study tested whether seeing the speaker's lip movements has a positive effect on SI. Any confounding effects were avoided by limiting visual input to the speaker's lip movements. Moreover, an innovative method – namely, measuring the duration of silent pauses in the target speeches – was used to capture the differences in cognitive load when the interpreter sees (or does not see) the speaker's lip movements.

The duration of silent pauses in the source speeches was indeed affected by whether or not the participants could see the speaker's lip movements: silent pause duration was significantly shorter when the speaker's lip movements were visible. This finding confirms the prediction of Seeber's (2017) cognitive load model for visual input in SI that the visibility of lip movements facilitates speech perception. Silent pause duration also depended on whether or not much noise was overlaid on the source speech, and to what extent. Provided that the SNR was sufficiently low, background noise led to longer silent pause durations. This is in line with Gerver's (1974) finding that noise increases the number of errors and omissions in SI and suggests that silent pause duration is a valid indicator for measuring cognitive load in SI.

To the best of my knowledge, this study is among the first to investigate the speaker's lip movements systematically as one type of visual input in SI. Further systematic studies on other types of visual input, such as the speaker's gestures and slides, could contribute to testing the predictions of Seeber's (2017) cognitive load model for visual input in SI. They could also contribute to obtaining a fuller understanding of the way in which different types of visual input affect the interpreter.

Acknowledgments

The study reported here is part of my PhD project. I am indebted to Professors Silvia Hansen-Schirra, Jukka Hyönä, Dörte Andres and Michaela Albl-Mikasa for their helpful advice and support in conducting and documenting the experiment.

References

Ahrens, B. (2004). *Prosodie beim Simultandolmetschen*. Frankfurt am Main: Peter Lang.

Ahrens, B. (2005). Prosodic phenomena in simultaneous interpreting: A conceptual approach and its practical application. *Interpreting 7* (1), 51–76. https://doi.org/10.1075/intp.7.1.04ahr

Albl-Mikasa, M. (2010). Global English and English as a lingua franca (ELF): Implications for the interpreting profession. *Trans-kom 3* (3), 126–148.

Anderson, L. (1994). Simultaneous interpretation: Contextual and translation aspects. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation*. Amsterdam: John Benjamins, 101–248.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67 (1). https://doi.org/10.18637/jss.v067.i01

Bernstein, L. E., Auer, E. T. & Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Communication 44* (1–4), 5–18. https://doi.org/10.1016/j.specom.2004.10.011

Benoit, C., Mohammadi, T., & Kandel, S. (1994). Effects of Phonetic Context on Audio-Visual Intelligibility of French. Journal of Speech and Hearing Research 37 (5), 1195–1203. https://doi.org/10.1044/jshr.3705.1195

Brancazio, L., Best, C. T. & Fowler, C. A. (2006). Visual influences on perception of speech and nonspeech vocal-tract events. *Language and Speech 49* (1), 21–53. https://doi.org/10.1177/00238309060490010301

Bühler, H. (1985). Conference interpreting: A multichannel communication phenomenon. *Meta 30* (1), 49–54. https://doi.org/10.7202/002176ar

Bühler, H. (1986). Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua 5* (4), 231–235.

Cecot, M. (2001). Pauses in simultaneous interpretation: A contrastive analysis of professional interpreters' performances. *The Interpreters' Newsletter 11*, 63–85.

Chernov, G. V. (1994). Message redundancy and message anticipation in simultaneous interpreting. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation*. Amsterdam: John Benjamins, 139–153. https://doi.org/10.1075/btl.3.13che

Chiaro, D. & Nocella, G. (2004). Interpreters' perception of linguistic and non-linguistic factors affecting quality: A survey through the World Wide Web. *Meta 49* (2), 278–293. https://doi.org/10.7202/009351ar

Chmiel, A., Szarkowska, A., Koržinek, D., Lijewska, A., Dutka, Ł., Brocki, Ł. & Marasek, K. (2017). Ear–voice span and pauses in intra- and interlingual respeaking: An exploratory study into temporal aspects of the respeaking process. *Applied Psycholinguistics 38* (5), 1201–1227. https://doi.org/10.1017/S0142716417000108

Davies, M. (2008-). Word frequency data. Retrieved from The Corpus of Contemporary American English (COCA): https://www.english-corpora.org/coca/ (accessed 19 March 2021).

Euopean Commission. (2009a). United Airlines rewards fittest people. Retrieved 4 February 2021, from Speech Repository website: https://webgate.ec.europa.eu/sr/speech/united-airlines-rewards-fittest-people

European Commission. (2009b). Disenchantment at work. Retrieved 4 February 2021, from Speech Repository website: https://webgate.ec.europa.eu/sr/speech/disenchantment-work

European Commission. (2012a). Demographic shift in Europe. Retrieved 4 February 2021, from Speech Repository website: https://webgate.ec.europa.eu/sr/speech/demographic-shift-europe

European Commission. (2012b). Greece in the doldrums. Retrieved 4 February 2021, from Speech Repository website: https://webgate.ec.europa.eu/sr/speech/greece-doldrums

Fox, J., & Weisberg, S. (2018). Visualizing Fit and Lack of Fit in Complex Regression Models with Predictor Effect Plots and Partial Residuals. Journal of Statistical Software 87 (9), 1-27. https://doi.org/10.18637/jss.v087.i09

Gerver, D. (1974). The effects of noise on the performance of simultaneous interpreters: Accuracy of performance. *Acta Psychologica 38* (3), 159–167. https://doi.org/10.1016/0001-6918(74)90031-6

Gerver, D. (1975). A psychological approach to simultaneous interpretation. *Meta 20* (2), 119–128. https://doi.org/10.7202/002885ar

Gerver, D. (2002). The effects of source language presentation rate on the performance of simultaneous conference interpreters. In F. Pöchhacker & M. Shlesinger (Eds.), *The interpreting studies reader*. London/New York: Routledge, 53–66.

Gieshoff, A. C. (2018). *The impact of audio-visual speech on work-load in simultaneous interpreting.* Doctoral thesis, University of Mainz.

Gile, D. (2009). *Basic concepts and models for interpreter and translator training. Revised edition*. Amsterdam: John Benjamins. https://doi.org/10.1075/btl.8

Goldman-Eisler, F. (1958). Speech analysis and mental processes. *Language and Speech 1* (1), 59–75.

Goldman-Eisler, F. (1961). The distribution of pause durations in speech. *Language and Speech 4* (4), 232–237.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. London/New York: Academic Press.

Goldman-Eisler, F. (2002). Segmentation of input in simultaneous translation. In F. Pöchhacker & M. Shlesinger (Eds.), *The interpreting studies reader*. London/New York: Routledge, 69–76.

I-hsin, I. L., Feng-lan, A. C. & Feng-lan, K. (2013). The impact of non-native accented English on rendition accuracy in simultaneous interpreting. *Translation & Interpreting 5* (2), 30–44.

Kramer, S. E., Kapteyn, T. S., Festen, J. M. & Kuik, D. J. (1997). Assessing aspects of auditory handicap by means of pupil dilation. *Audiology 36*, 155–164.

Lin, Y., Lv, Q. & Liang, J. (2018). Predicting fluency with language proficiency, working memory, and directionality in simultaneous interpreting. *Frontiers in Psychology 9*: 1543. https://doi.org/10.3389/fpsyg.2018.01543

Lo, S. & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology 6*: 1171. https://doi.org/10.3389/fpsyg.2015.01171

Macleod, A. & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology 21* (2), 131–141. https://doi.org/10.3109/03005368709077786

Massaro, D. W. & Cohen, M. M. (1999). Speech perception in perceivers with hearing loss: Synergy of multiple modalities. *Journal of Speech, Language, and Hearing Research 42* (1), 21–41. https://doi.org/10.1044/jslhr.4201.21

Mattys, S. L. & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language 65* (2), 145–160. https://doi.org/10.1016/j.jml.2011.04.004

Mattys, S. L., Brooks, J. & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology 59* (1), 203–243. https://doi.org/doi:10.1016/j.cogpsych.2009.04.001

McAllister, R. (2000). Perceptual foreign accent and its relevance for simultaneous interpreting. In B. Englund Dimitrova & K. Hyltenstam (Eds.), *Language processing and simultaneous interpreting: Interdisciplinary perspectives*. Amsterdam: John Benjamins, 45–63. https://doi.org/10.1075/btl.40.05mca

Mizuno, A. (2005). Process model for simultaneous interpreting and working memory. *Meta 50* (2), 739–752. https://doi.org/10.7202/011015ar

Moser, B. (1978). Simultaneous interpretation: A hypothetical model and its practical application. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication*. New York: Plenum Press, 353–368.

Moser-Mercer, B. (2003). Remote interpreting: Assessment of human factors and performance parameters. *Communicate! AIIC Webzine* (Summer 2003).

https://aiic.org/document/516/AIICWebzine_Summer2003_3_MOSER-MERCER_Remote_interpreting_Assessment_of_human_factors_and_performance_parameters_Original.pdf (accessed 16 March 2020).

Moser-Mercer, B. (2005). Remote interpreting: The crucial role of presence. *VALS-ASLA 81*, 73–97.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. Journal of Neuroscience Methods 162(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Plevoets, K. & Defrancq, B. (2016). The effect of informational load on disfluencies in interpreting: A corpus-based regression analysis. *Translation and Interpreting Studies 11* (2), 202–224. https://doi.org/10.1075/tis.11.2.04ple

Plevoets, K. & Defrancq, B. (2018). The cognitive load of interpreters in the European Parliament. *Interpreting 20* (1), 1–28. https://doi.org/10.1075/intp.00001.ple

Pöchhacker, F. (2005). From operation to action: Process-orientation in interpreting studies. *Meta 50* (2), 682–695. https://doi.org/10.7202/011011ar

Poyatos, F. (1984). The multichannel reality of discourse: Language-paralanguage-kinesics and the totality of communicative systems. *Language Sciences 6* (2), 307–337. https://doi.org/10.1016/S0388-0001(84)80022-4

Pradas Macías, M. (2006). Probing quality criteria in simultaneous interpreting: The role of silent pauses in fluency. *Interpreting 8* (1), 25–43. https://doi.org/10.1075/intp.8.1.03pra

Rackow, J. (2013). Dolmetschen als Kommunikation: Verbale und nonverbale Informationsverarbeitung im Dolmetschprozess. In D. Andres, M. Behr & M. Dingfelder Stone (Eds.), *Dolmetschmodelle – erfasst, erläutert, erweitert*. Frankfurt am Main: Peter Lang, 129–152.

Rennert, S. (2008). Visual input in simultaneous interpreting. *Meta 53* (1), 204–217. https://doi.org/10.7202/017983ar

Rennert, S. (2019). *Redeflüssigkeit und Dolmetschqualität: Wirkung und Bewertung*. Tübingen: Narr.

Roziner, I. & Shlesinger, M. (2010). Much ado about something remote: Stress and performance in remote interpreting. *Interpreting 12* (2), 214–247. https://doi.org/10.1075/intp.12.2.05roz

Sabatini, E. (2000). Listening comprehension, shadowing and simultaneous interpretation of two ‘non-standard’ English speeches. *Interpreting 5* (1), 25–48. https://doi.org/10.1075/intp.5.1.03sab

Seeber, K. G. (2011). Cognitive load in simultaneous interpreting. *Interpreting 13* (2), 176–204.

Seeber, K. G. (2015). Cognitive load in simultaneous interpreting: Measures and methods. In M. Ehrensberger-Dow, S. Göpferich & S. O’Brien (Eds.), *Interdisciplinarity in translation and interpreting process research*. Amsterdam: John Benjamins, 18–33. https://doi.org/10.1075/bct.72.03see

Seeber, K. G. (2017). Multimodal processing in simultaneous interpreting. In J. W. Schwieter & A. Ferreira (Eds.), *The handbook of translation and cognition* (pp. 461–475). Hoboken: John Wiley & Sons, 461–475.

Setton, R. (1999). *Simultaneous interpretation: A cognitive–pragmatic analysis*. Amsterdam: John Benjamins.

Seubert, S. (2017). Simultaneous interpreting is a whole-person process: Zur Verarbeitung visueller Informationen beim Simultandolmetschen. In M. Behr & S. Seubert (Eds.), *Education is a whole-person process: Von ganzheitlicher Lehre, Dolmetschforschung und anderen Dingen*. Berlin: Frank & Timme, 271–303.

Thomas, S. M. & Jordan, T. R. (2004). Contributions of oral and extraoral facial movement to visual and audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance 30* (5), 873–888. https://doi.org/10.1037/0096-1523.30.5.873

Tissi, B. (2000). Silent pauses and disfluencies in simultaneous interpretation: A descriptive analysis. *The Interpreters’ Newsletter 10*, 103–128.

Vatikiotis-Bateson, E., Eigsti, I.-M., Yano, S. & Munhall, K. G. (1998). Eye movement of perceivers during audiovisual speech perception. *Perception & Psychophysics 60* (6), 926–940. https://doi.org/10.3758/BF03211929

von Kriegstein, K., Dogan, Ö., Grüter, M., Giraud, A.-L., Kell, C. A., Grüter, T., Kleinschmidt, A. & Kiebel, S. J. (2008). Simulation of talking faces in the human brain improves auditory speech recognition. *Proceedings of the National Academy of Sciences 105* (18), 6747–6752. https://doi.org/10.1073/pnas.0710826105

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. New York: Springer.

Yu, W. & van Heuven, V. J. (2017). Predicting judged fluency of consecutive interpreting from acoustic measures: Potential for automatic assessment and pedagogic implications. *Interpreting 19* (1), 47–68. https://doi.org/10.1075/intp.19.1.03yu

Zwischenberger, C. (2010). Quality criteria in simultaneous interpreting: An international vs. a national view. *The Interpreters' Newsletter 15*, 127–142.

Address for correspondence

Anne Catherine Gieshoff
Zurich University of Applied Sciences
Theaterstrasse 15c
8401 Winterthur
Switzerland

annecatherine.gieshoff@zhaw.ch

Biographical note

Anne Catherine Gieshoff received her PhD in Interpreting Studies from the University of Mainz and currently holds a post-doc position in the interdisciplinary SNSF Synergia project CLINT – Cognitive load in Interpreting and Translating at ZHAW Zurich University of Applied Sciences. She is a member of the International Association for Translation and Intercultural Studies (IATIS) and the European Society for Translation Studies (EST). Her research focuses on cognitive load and quantitative and psychophysiological methods in conference interpreting.

Date received: 27 January 2019
Date accepted: 12 January 2021