Multimodal processing in simultaneous interpreting with text

Interpreters focus more on the visual than the auditory modality

Agnieszka Chmiel¹, Przemysław Janikowski² and Agnieszka Lijewska¹
¹ Adam Mickiewicz University in Poznań | ² University of Silesia in Katowice

The present study focuses on (in)congruence of input between the visual and the auditory modality in simultaneous interpreting with text. We asked twenty-four professional conference interpreters to simultaneously interpret an aurally and visually presented text with controlled incongruences in three categories (numbers, names and control words), while measuring interpreting accuracy and eye movements. The results provide evidence for the dominance of the visual modality, which goes against the professional standard of following the auditory modality in the case of incongruence. Numbers enjoyed the greatest accuracy across conditions possibly due to simple cross-language semantic mappings. We found no evidence for a facilitation effect for congruent items, and identified an impeding effect of the presence of the visual text for incongruent items. These results might be interpreted either as evidence for the Colavita effect (in which visual stimuli take precedence over auditory ones) or as strategic behaviour applied by professional interpreters to avoid risk.

Keywords: simultaneous interpreting, multimodal processing, multisensory processing, incongruence, visual modality

1. Introduction

Simultaneous interpreting is sometimes likened to the job of an air traffic controller working at a very busy airport (Zeier 1997; Leeson 2005). Just like the air traffic controller has to monitor the position, speed and altitude of numerous aircraft in the assigned airspace and communicate with pilots, the interpreter has
to juggle multiple tasks, such as listening to the source language text, expressing the same meaning in the target language and monitoring their own output. Interpreters’ performance oftentimes sparks awe among their listeners who admire bilingual processing under extreme temporal constraints. Although simultaneous interpreting seems a very difficult task in itself due to the cross-linguistic multitasking involved, professional interpreters are successful even when yet more multitasking is required, as is the case in simultaneous interpreting with text. In this type of interpreting, interpreters have access to the text of the speech to be read out by the speaker. In this case, the interpreter has to control three different channels – the speech produced by the speaker in the source language in the auditory channel, the self-produced interpreting in the target language in the auditory channel, and the source language text provided in the visual channel. Since the speaker may depart from the written text while delivering the to-be-interpreted speech, the interpreter has to constantly monitor both input channels. Additionally, self-monitoring gains importance in this type of interpreting because interpreters have to avoid source language interference not only from the auditory, but also from the visual channel (i.e., the written text).

Simultaneous interpreting with text is thus a real-life task that makes it possible to explore cross-linguistic and cross-modal processing. This paper presents an exploratory study aiming to examine how interpreters, as experts in processing multisensory input, cope with multimodal processing in simultaneous interpreting with text and, more specifically, how they deal with stimuli that are congruent and incongruent across modalities when performing simultaneous interpreting with text; in other words, when the written text matches the audio input to be interpreted or not. Simultaneous interpreters are a particularly interesting group for investigating cross-modal processing in a real-life task since their professional standards require them to give preference to the auditory input even if they have access to the transcript of the speech.

Interpreting Studies researchers are aware of the potential supporting role of the transcript in the process of simultaneous interpreting, especially when there is full congruence between what is said and what is written. Because of that they discourage looking away and conducting “pure SI” as an overall strategy (Gile 1997, 204). Yet, in the face of potential discrepancies between the two sources, both the theoretical and the professional standard seems to be uniform: “authoritative input still arrives through the acoustic channel” (Pöchhacker 2004, 19) and therefore interpreters should give priority to it (Jiménez Ivars 1999, 70). Practical applications of this rule of thumb (especially didactic) may differ. Some would-be interpreters are “instructed to take their eyes off the paper the second they notice that the speaker is no longer following the written text” (Schweda Nicholson 1989, 172); others are advised just to give “priority in attention to listening over
reading” (Setton and Dawrant 2016, 281). There seems to be a growing recognition of the need to practice this skill separately within the interpreter training curriculum (Schweda Nicholson 1989; Cammoun et al. 2009; Ivanov, Davies, and Naimushin 2014); however, the standard itself seems to remain unchallenged.

Such a standard goes against the commonly attested cognitive predisposition of humans to give preference to the visual rather than the auditory channel in the case of conflict (Colavita 1974; Koppen and Spence 2007; Sinnett, Spence, and Soto-Faraco 2007; Spence, Parise, and Chen 2012; Hirst, Cragg, and Allen 2018). The present study is one of the first to focus on congruence in multisensory processing in the task of simultaneous interpreting with text. This is also the first study in which three different types of incongruent stimuli (numbers, proper names and control words) in visual-auditory processing in interpreting are examined and in which cognitive accounts of cross-modal processing are applied. The following sections will briefly outline existing research on multimodal processing, including that in simultaneous interpreting. We will then describe the study conducted on conference interpreters asked to simultaneously interpret experimentally manipulated texts.

2. Existing research on multimodal processing

2.1 Multimodal processing in simultaneous interpreting with text

There are currently two theoretical accounts for simultaneous interpreting with text: the Effort Model (Gile 2009, 2018) and the Cognitive Load Model (Seeber 2017). However, both these models focus more on the demands created by the very presence of the written text during simultaneous interpreting than on the difficulties posed by possible mismatches between the text and the spoken input. According to Gile (2009), access to the written text is a facilitating factor since it reduces memory problems and acoustic difficulties, and helps comprehend unusual accents. Thus, the Reading Effort and the Listening Effort have a joint facilitatory effect. However, he also admits that the presence of the transcript may increase the “cognitive load arising from the need to follow both the vocal speech and the written text” (Gile 2009, 182).

According to Seeber’s conflict matrix (2017, 469–472), the total interference score (calculated as the sum of demand vectors and conflict coefficients for all overlapping modalities) is 11.6 for simultaneous interpreting and 14.8 for simultaneous interpreting with text, which indicates that the presence of text makes interpreting more (rather than less) demanding. This account is in line with other theoretical conjectures reviewed by Ivanov, Davies, and Naimushin (2014, 41),
who observe that the majority of scholars point out the added difficulty of having a text in the booth due to dual input (aural and visual), negative interference from the source text, and a risk of missing the speaker’s digressions from the written text.

However, the few empirical studies available actually provide evidence that is at variance with the stance of Seeber (2017), which is also supported by the review in Ivanov, Davies, and Naimushin (2014). These studies show a facilitating effect of the presence of the written text. Lambert (2004) found an overall facilitative role of written text as reflected in the significantly higher performance scores in simultaneous interpreting with text as compared to simultaneous interpreting without text. This result might be skewed by the fact that the participants were given ten minutes of preparation time before the actual interpreting in the condition with text. However, Spychała (2015) achieved similar results in her study that did not include preparation: she reported higher accuracy and quality of interpretation in the simultaneous interpreting with text as compared to regular simultaneous interpreting. In a more narrowly focused study, Korpal and Stachowiak (2015) found that access to visual information facilitated interpreting of numbers but they reported no correlation between viewing times and accuracy. Desmet, Vandierendonck, and Defrancq (2018) found the accuracy of interpreting numbers to increase by 30 percentage points on average when their group of interpreting students was provided with a mock-up technological support system that ‘extracted’ and presented the numbers visually in isolation and with a minimal delay. These results are further backed by survey-based research conducted with fifty interpreters, which concludes that “in simultaneous interpretation, the text is always a friend” (Cammoun et al. 2009, 130). Lamberger-Felber and Schneider (2009) focused on one particular aspect of multimodal processing in simultaneous interpreting and measured whether the availability of text would negatively impact performance by leading to more interference. They found no evidence for a negative influence as there was no difference in the occurrence of interference when comparing simultaneous interpreting with and without text.

All the above experimental studies used texts that were congruent with the speech stream. However, as has been mentioned above, the reality of authentic interpreted presentations is altogether different, as shown by Messina (1998, 154). She discovered a range of 3.1 to 14.7 additions and changes to the transcript per minute in texts produced by native speakers, and 1.01 to 3.7 per minute by non-native speakers at interpreted conferences. To date, however, there is only a single study focusing on the influence of congruent or incongruent visual information on performance in simultaneous interpreting. Korpal and Stachowiak (2015) provided their interpreters with visual input in the form of slides containing numbers congruent or incongruent with the auditory input. In general, they found
no facilitatory effect of the presence of visuals for professionals and decreased performance by trainees when confronted with incongruent information. Interpreters viewed the numbers similarly regardless of whether they were congruent or incongruent with the auditory input, while trainees looked less at the incongruent items. A potential explanation for this finding is the greater skill of professional interpreters in dealing with multiple incongruent inputs.

In summary, according to the theoretical accounts and exploratory research (Ivanov, Davies, and Naimushin 2014; Seeber 2017; Gile 2018), access to the visual text in simultaneous interpreting may make this task more difficult or comparable to regular simultaneous interpreting at best, due to multisensory input and visual interference. Experimental studies paint a different picture: interference does not increase (Lamberger-Felber and Schneider 2009) while accuracy does (Lambert 2004; Spychała 2015; Korpal and Stachowiak 2017) when the text is available. The facilitatory effect of the synergy between production and listening is also acknowledged in Gile’s model (2009). So far, only one study manipulated congruence in a task involving simultaneous interpreting with text and found no problems with integrating incongruent multimodal input on the part of professional interpreters (Korpal and Stachowiak 2015). Our study will thus extend this line of research to explore to what extent the written text might be of help or of hindrance in simultaneous interpreting.

2.2 Cognitive aspects of multimodal processing

As mentioned above, simultaneous interpreting with text involves processing input across two different sensory channels, and professional guidelines direct interpreters to attend to the auditory channel rather than to the visual input when the two are in conflict. In contrast, existing research on cognitive aspects of multisensory processing in adult humans has shown that participants generally display visually dominant behaviour in tasks where cross-modal stimuli are presented (for review see Spence, Parise, and Chen 2012).

Studies comparing the processing of simultaneously presented audiovisual stimuli show that participants fail to respond to auditory stimuli more often than to visual stimuli (Koppen, Alsius, and Spence 2008). Such visual dominance has been repeatedly observed in the so-called Colavita tasks where participants respond with key presses to unimodal targets (light flashes or beeps) and bimodal stimuli (light flashes and beeps presented concurrently) (Colavita 1974; Koppen, Alsius, and Spence 2008; Spence 2009). However, the Colavita visual dominance effect has also been observed in experiments involving more complex stimuli, including speech (Sinnett, Spence, and Soto-Faraco 2007; Koppen, Alsius, and Spence 2008; Sinnett, Soto-Faraco, and Spence 2008; Wille and Ebersbach 2016).
The effect is not modulated by stimulus intensity or response demands. Participants demonstrate it even if the auditory stimulus is twice the intensity of the visual one and regardless of the type of reaction required in the experiments. More importantly, the Colavita visual dominance effect cannot be reversed by directing attention to the audio channel. When participants were explicitly asked to attend to the audio channel, they still exhibited either visually dominant behaviour or no clear pattern of dominance (for review see Spence, Parise, and Chen 2012). The Colavita effect has been explained either by attention bias towards the visual modality, asymmetrical cross-modal accessory stimulus effects, or rapid forgetting of the auditory stimulus (Spence 2009). Regardless of its origins, if visual dominance overrides the professional guidelines for interpreting (i.e., if interpreters tend to interpret the written text and ignore the auditorily presented one in the case of incongruence), the Colavita effect might be one of the potential explanations of such a result in the current experimental study.

In this context, it is interesting to consider what happens in cross-modal processing when the stimuli are congruent, or, in other words, redundant across modalities. Matusz et al. (2015, 157) claim that “multiple sources of congruent information are integrated into a unified multi-sensory percut that triggers enhanced behavioural and/or neural responses.” Such an account is supported by a study by Laurienti et al. (2004), who used a feature discrimination task with congruent or incongruent stimuli presented in a cross-modal paradigm. The participants were asked to attend to both streams of information. Laurienti et al. (2004) found speeded responses for semantically congruent stimuli and slower responses for incongruent stimuli. A similar facilitation of bimodal presentation of congruent stimuli was found in other studies (e.g., Stubblefield et al. 2013, but see Koppen, Alsius, and Spence 2008).

Performance benefits for cross-modal congruent stimuli might result from the integration of cross-modal cues. Such stimuli might be processed more efficiently because they provide complementary information, which results in neural integration and ensuing facilitated processing (Doehrmann and Naumer 2008). This might be viewed as the cognitive foundation for Gile’s (2009) claims about cooperation and the facilitatory effect of the production and listening effort in simultaneous interpreting with text (see Section 2.1). Interestingly enough, even though existing research has indicated that congruence enhances performance, visual dominance in adult participants tends to be immune to the effects of semantic congruence. In other words, adults display visually dominant behaviour in both semantically congruent and incongruent trials (Koppen, Alsius, and Spence 2008; Wille and Ebersbach 2016).

It thus appears that congruent multisensory stimuli are integrated to enhance their processing while incongruent stimuli require greater cognitive control
(Doehrmann and Naumer 2008) to resolve the conflict. Visual dominance effects, though, tend to be consistently observed even when semantic congruence is manipulated in the task. However, one has to bear in mind that the majority of cognitive studies mentioned above made use of isolated stimuli and relatively unnatural tasks (e.g., button presses in response to light flashes and beeps), while our study examines an authentic and thus much more complex interpreting task. This makes the application of cognitive accounts to the interpreting task much less straightforward.

3. The present study

The purpose of the present study is to examine how conference interpreters cope with multisensory input in simultaneous interpreting with text and, more precisely, whether any facilitatory or impeding effects of the written text are evident when interpreters are faced with stimuli that are congruent or incongruent across the auditory and the visual input channels. A group of professional conference interpreters was asked to simultaneously interpret a spoken text from English (their L2) to Polish (their L1) while having access to the corresponding written text. As compared with the text delivered by the speaker, the written text included incongruences in proper names, numerical data and lexical items. Our dependent variables are interpretation accuracy and viewing times derived from eye-tracking data. As regards accuracy, the manipulation we have introduced is based on semantic congruence and so we might expect more accurate interpretations of cross-modally presented congruent items and less accurate interpretations of incongruent items in line with Laurienti et al. (2004); Doehrmann and Naumer (2008) and Gile’s theoretical account (2009).

Based on the nature of specific item types employed in the present study (numbers, proper names and control words) we expect differences in how they will be processed. Both numbers and names constitute non-contextual information in interpreting (Chmiel 2015). Since they are difficult to infer from the context, the interpreter has to rely on direct methods of transfer. Thus, they are likely to be more difficult to process and are expected to generate lower accuracy rates and longer viewing times than lexical items employed as control items. Since numbers are processed differently to letters (Dehaene and Cohen 2007) and are a common problem trigger in interpreting (Braun and Clarici 1996; Pinochi 2009), we expect that these items, of the three types of items used in the study, will be processed with the lowest accuracy and might trigger the longest viewing times, especially in the incongruent condition.
As far as the eye-tracking data are concerned, we expect a condition effect to modulate viewing times. If the viewing times of incongruent items are shorter than those of congruent items, we might conclude that the interpreters notice the conflict between mismatched stimuli and look away to avoid visual interference. However, if they are longer, this might be caused either by conflict resolution (Deutsch and Bentin 2001; see also Sullivan et al. 2016) or might be a strategic choice: interpreters decide to rely on the written text as a risk-avoidance strategy (Pym 2004) when they cannot keep up with the auditory input.

3.1 Participants

The participants were twenty-four freelance conference interpreters (thirteen female, eleven male), with an average age of thirty-eight ($SD=8.37$) and average experience of thirteen years ($SD=8.12$). The average number of working days per month was 5.94 ($SD=4.2$). In order to be included in the experimental group, the participants had to have at least four years of professional conference interpreting experience and to be active on the Polish conference interpreting market. They were recruited through a translation agency and compensated for their participation in the experiment. Their working language A (L1) was Polish and their B language (L2) was English. Some interpreters had additional C languages (L3s), such as French, German, Italian, Russian and Spanish. Their English proficiency was high, as their mean LexTale score (Lemhöfer and Broersma 2012) was 89.31 ($SD=9.31$). According to self-reported data, they spent 50% of their working time on interpreting into their A language (Polish). They all signed an informed consent form prior to the experiment.

3.2 Materials

The speech used for the simultaneous interpreting task was based on an authentic speech delivered in 2014 by Phil Hogan, an Irish nationality European Commissioner for agriculture. In the auditory modality, a recording of the speech was used. It was performed by an Irish native speaker of English with an Irish accent to encourage interpreters to look at the text. An accent more familiar to the participants would not pose an additional challenge in comprehension and the participants may have relied less on the written text.

The manipulated speech used in the present study was eleven minutes long and was recorded at a speaking pace of 115 words per minute. Sixty items evenly distributed across the text were selected as stimuli. These included twenty surnames of Irish individuals or names of places mentioned in the text, twenty numbers with various orders of magnitude (e.g., ‘1960s’, ‘30%’, ‘60’, ‘167 thousand’, ‘700
million’, ‘1.2 billion’) and twenty control words which were neither proper names nor numbers (they were content words in sentences – either nouns, verbs or adjectives). Incongruent versions of the stimuli were prepared as follows: incongruent names were other authentic surnames occurring in Ireland, incongruent numbers were numbers with changed digits but retained order of magnitude, incongruent words were words from the same part of speech category, matched for length and frequency, plausible as replacements of the original words in the sentences. The mean cross-linguistic similarity of lexical items and their translations measured by the normalised Levenshtein distance was 0.72 (ranging from 0.2 to 1.0, $SD = 0.21$).

Great care was taken to match congruent and incongruent stimuli because they constitute the basis for our comparison of the congruent and incongruent condition. The use of the target item pairs allowed us to manipulate the level of cross-modal congruence and to hold all other linguistic properties of the to-be-translated text constant.

Two versions of the experimental text were further prepared for the visual modality. Each version included ten congruent and ten incongruent stimuli from each type (names, numbers, words). If a given stimulus was congruent in version A, it was incongruent in version B. Thus, the auditorily presented text was always the same while the participant saw either version A or version B of the written text. The written text was divided into nine equal parts to be displayed on nine screens during the experiment to make certain that the font was comfortable to read and to ensure eye-tracking data quality. It was displayed in Arial, font size 21pt on a 21-inch LSD monitor.

3.3 Procedure

The experiment was prepared and run in Experiment Builder (SR Research). The eye-tracking data were recorded by the EyeLink 1000 Plus eye-tracker. The participants sat in front of the monitor and listened to the speech recording via headphones. They were instructed to perform simultaneous interpreting. They were told that the text of the speech would be displayed on the screen in front of them. They were asked to adhere to professional standards, one of which is that, in case of discrepancy, the orally presented text takes precedence over the written text. However, this standard was not explicitly mentioned to the participants so as not to attract their attention to the possibility of having to deal with incongruences. The session started with a thirteen-point calibration procedure. Then instructions were presented both auditorily and visually. The instructions were voiced by the same speaker who had recorded the experimental speech so that the participants could get used to his accent. The instructions included an interpreter’s brief; in
other words, the speaker explained the context of the to-be-interpreted speech. When the interpreter finished interpreting the text on each of the nine screens, the experimenter changed the screen and the next part of the written text was displayed and the recording continued. The experimenter monitored eye-tracking data quality and if needed, recalibration was performed between screen changes. The experimental session lasted between fifteen and twenty minutes.

4. Results

Accuracy of interpreting was assessed for each of the experimental items. The renditions were compared to the source text items in the audio input. Since they were single words or numerals, decisions on accuracy were fairly clear. If the interpretation corresponded to what the interpreter heard, it was classified as accurate. If it did not, it was classified as inaccurate and further assigned to the visual interference category (if it corresponded to the incongruent item in the visual input) or the inaccurate interpretation category (which included omissions and inaccurate renditions other than ones based on the visual input). We are aware that omissions might stem from the interpreter’s strategic decision to cope with information overload. However, since our further analysis focused on the faithful renditions only and since it was impossible to unambiguously separate omissions resulting from a strategy from those due to error, we decided to include omissions in the inaccurate interpretation category. We then removed those items that did not receive any fixations. Figure 1 presents detailed accuracy results by stimulus type and congruence for stimuli that were viewed (according to the eye-tracking data).

We calculated skipping rates as a percentage of experimental items that received no fixation out of all experimental items. We analysed them per stimulus type and found the following skipping rates: 18% for names, 22% for numbers and 23% for control words. A chi-square test showed no significant difference in skipping rates between these types ($\chi^2 = 4.50, df = 2, p = .16$).

To examine the effects of condition and item type on accuracy, we fitted a general linear model with random intercepts for participants and items, and with fixed effects of condition (congruent, incongruent) and type (number, name, control word). We found a statistically significant effect for condition ($B = -3.34$, $SD = 0.17$, $z = -19.13, p < .001$). Congruent items (with mean accuracy rate of 86.69%, $SD = 0.34$) were interpreted more accurately than incongruent items (with mean accuracy rate of 24.51%, $SD = 0.43$). We also found an effect of type, with a statistically significant difference between names and numbers ($B = 0.70$, $SD = 0.30$, $z = 2.27, p < .05$) and between words and numbers ($B = 0.60$, $SD = 0.30$, 46 Agnieszka Chmiel, Przemysław Janikowski and Agnieszka Lijewska
The mean accuracy rate was the highest for numbers ($M=60.20\%, SD=0.49$), followed by words ($M=54.64\%, SD=0.49$) and names ($M=52.08\%, SD=0.50$).

We also found two interactions, one suggesting that condition differently modulates the accuracy of words and names ($B=-1.48$, $SD=0.35$, $z=-4.15$, $p<.001$) and one pointing to condition modulating the accuracy of words and numbers ($B=-1.71$, $SD=0.38$, $z=-4.46$, $p<.001$). As shown in Figure 2, incongruence hampers accuracy of control words less than accuracy of numbers and names. We estimated the model’s goodness of fit based on conditional and marginal $R^2$ measures (Nakagawa and Schielzeth 2013) using the sem.model.fits function from piecewiseSEM package in R (Lefcheck and Freckleton 2016): $R^2_{GLMM(c)}=.51$, $R^2_{GLMM(m)}=.37$.

We also analysed the accuracy in the incongruent condition, which included three categories of output: accurate interpretations, inaccurate interpretations, and inaccurate interpretations due to visual interference. A chi-square test shows significant differences in these categories of output for names, numbers and control words ($\chi^2=35.3$, $df=4$, $p<.001$, Cramer’s $V=.25$). The highest rate of visual interference was recorded for names (71% of all interpretations in the incongruent condition), followed by numbers (62%) and control words (43%).

We then analysed the eye-tracking data, specifically the total viewing duration for those experimental items that were not skipped. After visual inspection of the data, durations exceeding 3,000 ms were removed from further data analysis.
(approximately 1% of data removed). We further applied a log-transformation to the data to obtain a normal distribution. However, for clarity of presentation, we report non-transformed means below. We fitted a linear mixed model with fixed factors of condition, type and accuracy (and their interactions), and random intercepts for items and participants. The analysis revealed an effect of condition ($B = 0.17$, $SD = 0.06$, $t = 2.87$, $p < .01$), with incongruent items viewed longer ($M = 761$ ms, $SD = 560$) than congruent items ($M = 708$ ms, $SD = 507$). We also found an effect for accuracy ($B = 0.15$, $SD = 0.06$, $t = 2.39$, $p < .05$) with accurately interpreted items viewed longer ($M = 739$ ms, $SD = 538$) than inaccurately interpreted items ($M = 728$ ms, $SD = 530$). There was a condition by accuracy interaction ($B = -0.31$, $SD = 0.12$, $t = -2.44$, $p < .05$), which suggests shorter viewing of inaccurately interpreted congruent items. We found neither an effect for type nor an interaction of condition and type. The model’s goodness-of-fit measures (Nakagawa and Schielzeth 2013; Lefcheck and Freckleton 2016) are as follows: $R^2_{GLMM(c)} = .19$, $R^2_{GLMM(m)} = .03$. Table 1 shows total viewing durations across conditions and item types.

Finally, we also analysed skipping data to see whether looking at the experimental items influenced the accuracy of participants’ interpretations. We thus calculated a correlation between the participants’ accuracy rates and skipping
Table 1. Total viewing time of experimental items

<table>
<thead>
<tr>
<th>Condition</th>
<th>Type</th>
<th>Mean (in ms)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Names</td>
<td>630</td>
<td>449</td>
</tr>
<tr>
<td></td>
<td>Numbers</td>
<td>855</td>
<td>575</td>
</tr>
<tr>
<td></td>
<td>Control words</td>
<td>638</td>
<td>458</td>
</tr>
<tr>
<td>Congruent</td>
<td>Names</td>
<td>726</td>
<td>525</td>
</tr>
<tr>
<td></td>
<td>Numbers</td>
<td>764</td>
<td>547</td>
</tr>
<tr>
<td></td>
<td>Control words</td>
<td>796</td>
<td>608</td>
</tr>
<tr>
<td>Incongruent</td>
<td>Names</td>
<td>726</td>
<td>525</td>
</tr>
<tr>
<td></td>
<td>Numbers</td>
<td>764</td>
<td>547</td>
</tr>
<tr>
<td></td>
<td>Control words</td>
<td>796</td>
<td>608</td>
</tr>
</tbody>
</table>

rates (calculated as the percentage of items that received no fixations) for congruent and incongruent items separately. There was no correlation between skipping rate and accuracy in the congruent condition \(r = .05, p = .79\), but we found a strong positive correlation between skipping rate and accuracy \(r = .62, p < .01\) for incongruent items, suggesting that the more the participants skip (i.e., the less they look at the experimental items), the higher their accuracy scores (see Figure 3). This suggests that the presence of the written text (if used and viewed by the interpreter) does not influence accuracy for congruent items and decreases accuracy for incongruent items. We also calculated a correlation between mean viewing duration and accuracy rate, which turned out to be insignificant in the congruent condition \(r = -.19, p = .37\), but moderately negative and marginally significant in the incongruent condition \(r = -.40, p = .051\) (see Figure 4). This means that longer viewing durations of incongruent items lead to less accurate interpretations.

Figure 3. Correlation between skipping rate and accuracy in the incongruent condition
5. Discussion

The aim of this study was to examine how conference interpreters deal with multisensory input in the task of simultaneous interpreting with text. As professional interpreters frequently process information from various modalities, we asked them to simultaneously interpret a text that included incongruences between the audio version and the written version also made available to the participants during the experimental task. We wanted to determine if the presence of the written text would be helpful when processing congruent items (thanks to the integration of multisensory information) and harmful when processing incongruent items (due to the Colavita visual dominance effect).

In line with our expectations, interpreters coped with congruent items better and the accuracy rate of their interpretations was higher in the case of congruent items as compared to the incongruent ones, which might lend support to Laurienti et al. (2004) and Doehrmann and Naumer (2008). We found that in the case of bimodally presented incongruent stimuli the interpreters mostly focused on one modality only as the majority of inaccurate translations in the incongruent condition were due to visual interference (in 71% of cases for names, 62% for numbers, and 43% for control words). These findings might be explained as the consequence of either a strategic or an automatic response. First, interpreters might have chosen to rely on the text because they missed the information from the auditory input (Jiménez Ivars 1999; Sile 2003; Gile 2009). Since, as in any type of translation, interpreters strove to produce a failure-free output, they applied a risk-avoidance strategy (Pym 2004) and relied on the accessible written text.
when faced with difficulties in processing both inputs. Another potential explanation is related to the cognitive accounts of multisensory processing. It might be the case that visual stimuli took precedence over auditory ones, in line with the Colavita effect (Colavita 1974; Spence, Parise, and Chen 2012) without the interpreter’s conscious decision to focus on the visual input channel only.

Obviously the current design does not make it possible to differentiate between the cases in which visual interference in the interpreting of incongruent items was due to the Colavita effect and the cases in which the participants noticed the incongruence but strategically chose to base their interpretation on the visual modality as a risk-avoidance strategy (i.e., they could not retrieve the memory trace of the auditory stimulus and used the written one instead). More research is needed to disentangle these effects. Another limitation and an avenue for further research is the influence of the difficulty of audio input on visual interference in simultaneous interpreting with text. In our design we decided to only test the extreme situation of a highly demanding accent. It remains to be seen whether the observed patterns persist when speakers with less challenging accents are involved.

At a more fine-grained level, we manipulated the type of congruent and incongruent stimuli and predicted that, due to their specificity, numbers would be interpreted least accurately, followed by names and words. Contrary to our predictions, we found that accuracy was highest for numbers and differed statistically from that for names and control words. This seemingly counterintuitive result might be easily explained. Because numbers are notorious for causing trouble to interpreters (Gile 2009) interpreters may have allocated extra effort and attention to processing them. Additionally, numbers have a rather straightforward semantic mapping across languages. This one-to-one mapping across the applied language pair largely simplifies lexical selection, which might have effectively counteracted the extra processing demands in the congruent condition (Prior et al. 2015). In the incongruent condition the fast and unambiguous access to respective units in the mental lexicon may have provided the interpreters with more time for conflict resolution and thus reduced the negative effect of incongruence on accuracy.

We found an interesting condition by type interaction in the analysis of accuracy data, showing that incongruence lowered the accuracy of control words less than the accuracy of numbers. This interesting result could perhaps be partially explained by much higher context-dependence in the case of processing control words as compared to numbers and names. When the advantage of having a congruous visual backup is lost and when conflict is introduced instead, in the case of the control words the immediate sentence context may work to the benefit of the interpreter perhaps even at the level of strengthening the short-term memory trace (see Iodice et al. 2018 for the effects of context on recall). Numbers and
names are far less contextually inferable and thus could suffer more in the incongruous condition.

Our analysis of the eye-tracking data confirmed our predicted condition effect. The findings show that incongruent items were viewed longer than congruent ones. This means that when the conflict between the mismatched stimuli was noticed, the interpreters did not look away to avoid visual interference but focused more on the visual stimuli in an attempt to resolve the conflict between the modalities. Interestingly, accurately interpreted items across conditions were viewed longer than inaccurately interpreted items. This may simply testify to more superficial processing leading to erroneous choices. Somewhat adventurously we may posit that at least a part of these inaccuracies may have fallen within the range of a pre-semantic visual inspection (below 300–400 ms; Whiting, Shtyrov, and Marslen-Wilson 2015; see Table 1). This is not very hard to imagine, especially as we need to remember that most of the processing in our task happened under rather extreme cognitive load in which saving effort on any of the components involved was highly desirable. Thus, shortening the visual inspection of the input and limiting it to just formal aspects may have been a coping strategy producing the effect described above. This, however, is only a very tentative explanation that would require further experimental enquiry.

The correlations between accuracy and skipping rates for the two conditions might help us disentangle the positive and negative effects of having access to the written text in simultaneous interpreting. We found no correlation between accuracy and skipping rate in the congruent condition. This means that accuracy was not influenced by viewing or skipping experimental items. Thus, we found no evidence for the facilitatory effect of text availability when interpreting selected experimental items. This is at a variance with Spychała (2015) and Korpals and Stachowiak (2015). These differences might be due to the fact that Spychała’s study involved trainees and the one by Korpals and Stachowiak included only three professional interpreters as participants. For the incongruent condition, we found a strong positive correlation, suggesting that more skipping (or less viewing) leads to better accuracy. Thus, it might be concluded that access to the written text in simultaneous interpreting with text might have an impeding effect on performance when it is not fully congruent with the audio input as it entails more processing load (Gile 2009) due to conflict resolution.

So what do these results tell us about multimodal processing in simultaneous interpreting with text? Faced with excessive multimodal input, interpreters focus their attention on the dominant visual modality and fall into the trap of interpreting incorrect stimuli that are incongruent with the auditory text. As Gile (2009) points out, interpreters frequently work close to cognitive saturation level and
make errors due to overload. Our study shows that this is indeed the case when processing incongruent bimodally presented stimuli.

We found no facilitation effect, even in the congruent condition. This might be explained by the fact that “a multimodal facilitation effect during SI with text is contingent on the synchronicity of the two signals (e.g., text and speech)” while “asynchrony of signals on different channels is expected to increase cognitive load” (Seeber 2016). The current analysis did not allow us to determine to what extent these two signals were perceived simultaneously (in the second stage of the current research project, we will conduct further analyses of time lag measures such as eye-ear span, that is, the interval between hearing and viewing the experimental items; ear-voice span; and eye-voice span) (Janikowski, Chmiel, and Lijewska, in preparation). In fact, we might suspect that the interpreters did lag behind in viewing the text as compared to the auditory input and strategi- cally chose to follow the written text to reduce failure and mitigate risk (Pym 2004), since visual interference rates were high (over 70% for names and 60% for numbers) and viewing times were long in the incongruent condition. If the synchronous perception of both channels is confirmed by the ear-eye span, we might resort to explanations based on the cognitive multimodal processing theories and claim that the visual modality takes precedence over the auditory one in line with the Colavita visual dominance effect (Colavita 1974) despite professional standards, according to which interpreters should adhere to what the speaker says and not to the speech transcript.

6. Conclusions

Conference interpreters have to face multisensory input when performing the real-life task of simultaneous interpreting with text. Since simultaneous interpreting is a difficult task with high cognitive load, the addition of another input in a different modality has major, but not uniform, ramifications for the interpreters’ performance. The participants in our study achieved higher accuracy when interpreting congruent items and focused much on the written modality, which resulted in visual interference in the incongruent condition. Faced with information from multisensory sources, interpreters in this study tended to adhere to the written modality. This might be a strategic choice to avoid failure when information from the auditory modality was not processed sufficiently, or might be an automatic mechanism based on the Colavita visual dominance effect. Further studies are needed to ensure that the cross-modal stimuli are perceived synchronously by the participants in order to confirm that the Colavita mechanism is indeed at work in simultaneous interpreting with text.
In exploring correlations between viewing the text and accuracy, we found no facilitation effects for congruent items and a negative correlation for incongruent items. This means that we found no evidence for a facilitatory effect of the written text’s presence when interpreting congruent stimuli and evidence for a negative effect of the written text’s presence for incongruent stimuli – in line with the theoretical account by Seeber (2017) and partially in line with Gile’s (2009) model. Obviously, this study was limited to non-contextual stimuli (numbers and names) and control words. Further studies are needed to explore how text availability impacts syntax or overall quality of interpreting. In our study, numbers turned out to be the most accurately interpreted items in the congruent condition due to their simple cross-linguistic semantic mapping, while control words were the most accurately interpreted items in the incongruent condition as they benefited from sentential context. The bimodal mismatch resulted in longer viewing times as the participants tried to resolve the conflict between the auditory and the visual modality. In experimentally manipulating the task of simultaneous interpreting with text we hope to have shed some new light on the complexities of multisensory processing performed by experts who are used to coping with such bimodal input.

References


Address for correspondence

Agnieszka Chmiel
Department of Translation Studies
Faculty of English
Adam Mickiewicz University in Poznań
ul. Grunwaldzka 6
60-780 POZNAŃ
Poland
achmiel@amu.edu.pl
https://orcid.org/0000-0002-2138-3974

Co-author information

Przemysław Janikowski
University of Silesia in Katowice
przemyslaw.janikowski@us.edu.pl

Agnieszka Lijewska
Adam Mickiewicz University in Poznań
alijewska@wa.amu.edu.pl

Publication history

Date received: 19 December 2018
Date accepted: 18 December 2019
Published online: 21 January 2020