Mid-Air Hand Gestures for Post-Editing of Machine Translation

Rashad Albo Jamara[1], Nico Herbig[1], Antonio Krüger[1], Josef van Genabith[1,2]
[1]German Research Center for Artificial Intelligence (DFKI),
Saarland Informatics Campus, Germany
[2]Department of Language Science and Technology,
Saarland University, Germany
rashad.jamara@gmail.com
{nico.herbig, krueger, josef.van_genabith}@dfki.de

Abstract

To translate large volumes of text in a globally connected world, more and more translators are integrating machine translation (MT) and post-editing (PE) into their translation workflows to generate publishable quality translations. While this process has been shown to save time and reduce errors, the task of translation is changing from mostly text production from scratch to fixing errors within useful but partly incorrect MT output. This is affecting the interface design of translation tools, where better support for text editing tasks is required. Here, we present the first study that investigates the usefulness of mid-air hand gestures in combination with the keyboard (GK) for text editing in PE of MT. Guided by a gesture elicitation study with 14 freelance translators, we develop a prototype supporting mid-air hand gestures for cursor placement, text selection, deletion, and reordering. These gestures combined with the keyboard facilitate all editing types required for PE. An evaluation of the prototype shows that the average editing duration of GK is only slightly slower than the standard mouse and keyboard (MK), even though participants are very familiar with the latter, and relative novices to the former. Furthermore, the qualitative analysis shows positive attitudes towards hand gestures for PE, especially when manipulating single words.

1 Introduction

In a well-connected world, translation is of ever-increasing importance (Bassnett, 2013). To meet translation demands, machine translation (MT) is often employed as a cheaper and faster alternative to human translation (HT) (O'Brien, 2012). Even though MT has improved drastically over the last 5 years, discussions about reaching human parity are still ongoing (Läubli et al., 2020) and limited to a small set of language pairs and domains for which ample training data is available. For most application scenarios, however, MT quality is far from reaching the quality of highly trained professionals. In an attempt to combine the best of both worlds, post-editing (PE) is becoming common practice, where human translators use raw MT output and make the necessary changes to produce an acceptable level of quality (Koponen, 2016). Although translators have approached PE with fear and skepticism (Lagoudaki, 2009), more recent studies found that nowadays translators are more open to it and that much of the original dislike was attributed to outdated perceptions of MT quality (Plitt and Masselot, 2010; Green et al., 2013). Independent of translators' perceptions, studies found that PE increases productivity and decreases errors compared to translation from scratch (Green et al., 2013).

PE changes the translation task from mostly text generation to text editing, which involves an increased usage of navigation and deletion keys (Toral et al., 2018). As a result, translators need better support with text editing operations, which raises the question whether interaction modalities other than mouse and keyboard can be beneficial for PE. An interaction modality that has gained attention in other research areas (Koutsabasis and Vogiatzidakis, 2019) but so far remains unexplored for PE is **mid-air hand gestures**.

In this paper, we (i) investigate which mid-air gestures combined with the keyboard (GK) are suitable for which text-editing operations in PE, (ii) build a prototype supporting PE using GK, and (iii) analyze editing times and subjective feedback on mid-air hand gestures compared to mouse and keyboard (MK) for specific PE operations. To address these goals, we conducted a gesture elicitation study (GES) with professional translators, resulting in a set of gestures for different editing tasks, which were then implemented in a prototype.

Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, pages 6763–6773
August 1–6, 2021.

Our experiment shows that, surprisingly, editing durations for most PE tasks were very similar in the conditions GK and MK, even though participants were much more experienced with the latter. Furthermore, participants prefer manipulating *single items*[1] using gestures, while manipulating a *group of items*, which involves more complex text selection, received poorer subjective feedback.

2 Related Work

In this section, we present related research on translation environments, multi-modal approaches to PE, and mid-air gestures for text editing tasks.

2.1 CAT Tools and Post-Editing

In recent years, most translators use computer-aided translation (CAT) tools for translation (Coppers et al., 2018). CAT tools are workflow systems offering features like translation memory (TM), MT, or terminology management (Van den Bergh et al., 2015; Koskinen and Ruokonen, 2017). Translators prefer to use CAT tools as they enhance terminology consistency, increase productivity, and improve the general quality of translations (Rossi and Chevrot, 2019; Moorkens and O'Brien, 2017).

While TM is still often valued more than MT (Moorkens and O'Brien, 2017), a recent study by Vela et al. (2019) shows that professional translators who were given a choice between translation from scratch, TM, and MT, chose MT in 80% of cases, highlighting the importance of PE of MT. Apart from translators' preference, Toral et al. (2018) found that PE phrase-based and neural MT (PBSMT and NMT) output increased productivity by 18% and 36% respectively compared to HT.

PE also changes the interaction patterns compared to manual translation from scratch (Carl and Jensen, 2010), leading to a significantly reduced amount of mouse and keyboard events (Green et al., 2013). At the same time, navigational and deletion key usage increases by 72% during PE of NMT compared to HT (Toral et al., 2018). This motivates our decision to explore modalities other than MK for PE and to specifically focus on efficient navigation and deletion.

2.2 Multi-Modal Approaches

Previous studies already explored modalities other than MK: The CASMACAT tool (Alabau et al., 2014) allows users to hand-write text with an e-pen. Studies on mobile PE via touch and speech (O'Brien et al., 2014; Torres-Hostench et al., 2017) show that participants especially like reordering words through touch drag and drop, and prefer voice input when translating from scratch, but stick to the iPhone keyboard for small changes. Zapata (2016) also explores the use of voice- and touch-enabled devices; however, their study did not focus on PE, and used Microsoft Word instead of a proper CAT environment. Teixeira et al. (2019) explore a combination of touch and speech for translation from scratch, translation using TM, and translation using MT and found that their touch implementation received poor feedback, while dictation turned out to be quite useful.

We started our research on multi-modal CAT tools with an elicitation study (Herbig et al., 2019), which showed that pen, touch, and speech interaction, as well as combinations thereof, should be combined with mouse and keyboard to improve PE of MT. A prototype based on the proposed interactions allows users to "directly cross out or hand-write new text, drag and drop words for reordering, or use spoken commands to update the text in place" (Herbig et al., 2020b). Its evaluation with professional translators further showed that depending on the editing operation, different input modalities performed well (Herbig et al., 2020a).

To date, mid-air gestures have only been addressed in our elicitation study (Herbig et al., 2019), where participants did not expect them to be particularly useful. However, participants only considered gestures on their own (i.e. also for text entry), and thus the combination with the keyboard merits further investigation, both in terms of an elicitation study and even more so in a practical evaluation of a prototype.

2.3 Mid-Air Hand Gestures

Hand gestures provide an intuitive and natural way of interaction (Sharma and Verma, 2015; Ortega and Nigay, 2009), but the design of appropriate gestures depends on the application type and context (Wachs et al., 2011; Weichert et al., 2013; Nielsen et al., 2003). Gestures must be easy to learn and memorize, comfortable to perform, and should be metaphorically meaningful (Wachs et al., 2011; Weichert et al., 2013).

Ortega and Nigay (2009) explored the use of mid-air finger pointing to replace the mouse and showed that this approach significantly reduces the

[1] Item(s) refers to word(s) and/or punctuation mark(s).

switching time compared to MK (almost to zero). However, research on text editing using hand gestures is scarce. One exception is Rives et al. (2014), who presented the idea of using gestures to perform the operations cut, copy, paste, select, undo, and delete to edit a document using gestures. In their concept, the user enters the edit mode through a special gesture and then draws in the air to perform the above operations, e.g. a "X" for deletion.

To find a suitable and concise set of gestures for PE operations, we conduct a GES.

3 Gesture Elicitation Study

A GES is a form of participatory design (Morris et al., 2014) where users are incorporated in the design process to inform an appropriate gesture set for a given application. Important aspects include leading participants away from technical thinking (Nielsen et al., 2003), making them assume that gesture recognition is perfect, and considering their behavior as always acceptable (Wobbrock et al., 2009). They should only be informed about the essential details of the task to avoid bias towards particular approaches (Wobbrock et al., 2005).

We conduct a GES for three reasons. Firstly, there is no universal gesture set suitable for all applications (Nielsen et al., 2003). Secondly, users prefer gestures designed through elicitation studies, because professional designers tends to generate more physically and conceptually complex gestures (Morris et al., 2014). Thirdly, to the best of our knowledge, there is no other GES for text editing using GK which we could rely on.

In our GES, we employed the guessability approach (Wobbrock et al., 2005) which is intended to increase immediate usage of interfaces. It consists of three phases: (1) defining so-called referents (i.e. common operations) that should be achievable through the system, (2) asking participants to propose a gesture for each referent, and (3) analyzing the collected data to generate the final gesture set.

3.1 Method

Due to the COVID-19 pandemic we conducted an online GES. Prior to commencing the study, ethical clearance was sought from the university ethical review board. The study took 30 to 65 minutes per participant (avg: 46 minutes).

Participants: Fourteen right-handed freelance translators (with 14 different nationalities, 7 female and 7 male) were hired to participate in the study (avg age: 28, SD: 4.56). Years of professional experience ranged from 2 to 15 years (avg: 5.29, SD: 3.43), offering a total of 19 language pairs. In terms of CAT tool experience, about 2/3 of the participants reported using CAT tools to aid translation, with 1 to 4 years of experience. Overall, participants were often in the earlier stages of their professional careers. Three of the participants already had experience with gesture-based interfaces such as a TV remote control. However, they rated their level of experience with gestural interfaces as "Bad" to "Neutral".

Referents: Referents are described as the effect which is triggered by a gesture (Wobbrock et al., 2009). The referents used in elicitation studies are an essential part, since the results established are limited to this set. In our case, referents are PE operations; we will thus use referents and operations interchangeably. To find good referents, we looked at different PE task classifications discussed in the literature. Popovic et al. (2014) propose 5 PE operations: correcting word form, correcting word order, adding omission, deleting addition, and correcting lexical choice. Koponen (2012) additionally distinguishes between moving single words or groups of words and the distance of the movement. Based on these studies as well as our previous elicitation procedure (Herbig et al., 2019), we propose the referents presented below as PE tasks for which we explore gestural input.

- I: Insertion
- D_s: Deleting a single item
- D_g: Deleting a group of items
- RP_s: Replacing a single item
- RP_g: Replacing a group of items
- RO_s: Reordering a single item
- RO_g: Reordering a group of items

Performing those referents implicitly includes other operations, namely selecting a position, a word, or a group of words/characters.

Procedure: We interviewed each participant online via a video conferencing platform. The first part of the study introduced PE of MT, discussing the current use of mouse and keyboard in CAT tools, and presenting the idea of mid-air hand gestures for PE without showing any concrete gestures

that could induce bias. Participants were then asked to fill out an online questionnaire capturing their demographics as well as other questions concerning CAT tools and MT in general. They were also informed that they should assume perfect recognition and that all proposals are valid. After each gesture proposal, participants supplied subjective ratings on 7-point Likert scales (7 = "strongly agree") as to whether the gesture is: (a) a good match for its intended purpose, (b) easy to perform, and (c) a good alternative to MK. Additionally, we used a think-aloud protocol and videotaped the session for subsequent analysis. Our referents were counterbalanced to avoid systematic errors.

Analysis: For the analysis, we grouped similar gestures based on the number of hands involved, their physical attributes and movement direction. We report the largest groups per referent, but also the agreement rate (AR), "characterizing the level of consensus between participants' proposals elicited" (Vatavu and Wobbrock, 2015). A high AR suggests that the most frequent gesture proposal is guessable and intuitive. However, less frequent proposals can still yield interesting insights.

3.2 Results & Discussion

Unlike static gestures, dynamic gestures are hard to illustrate through images; therefore, we created a simple website that shows recorded animations of gestures for each participant and groups them based on the referent[2].

While analyzing the data, consistent patterns emerged: Similar to the way the mouse is used, participants performed all referents by first selecting the text, then performing the editing operations, e.g. deleting. Consequently, we decided in our analysis to separate the selection gestures from the editing operation gestures, analyzing and discussing each separately. In addition, the proposed selection gestures are divided into two types: the selection of a single item and the selection of a group of items.

Group Selection: 8 unique gestures were proposed for group selection for the referents D_g, RP_g, and RO_g[3], with the same AR of 0.13 for each. Two of these gestures were the most common, namely *both indices* (pointing with index fingers and moving them apart to select: see Figure 1a) and *index + thumb* (pointing with pinched index finger and thumb and separating them to select a range). *Both indices* was rated higher on "ease" than *index + thumb*, but received almost identical ratings for "good match" and "alternative", indicating a slight preference for using both index fingers. The remaining 6 proposals were interesting ideas like using a certain number of fingers to specify the number of words to select, however, none of these proposals reached agreement.

Single Item Selection: Participants proposed 5, 9, and 8 different gestures for the referents D_s, RP_s, and RO_s, respectively. Consequently, the high number of different proposals for replacing and reordering reduced the AR to 0.08 (RP_s) and 0.09 (RO_s) compared to 0.16 for D_s. Participants mostly proposed the same single item selection gesture for all subsequent referents, highlighting the importance of counter-balancing. However, the *index + thumb* and *both indices* appear to also be preferred in selecting a single item, but with slightly varying agreement scores compared to group selection. In addition, the gesture *pointing* (where a participant points with the index finger to place the cursor on the item) was highly preferred for single item selection. The *double-tap* gesture was also proposed 3, 2, and 1 times for the referents D_s, RO_s, and RP_s, respectively.

When asked about the reasons for their proposals, participants (p) gave responses such as p3: "It is easy and intuitive" or p5: "It is really easy to select the start and then slide it to select".

Editing Operations: Unlike selection gestures, editing operations received very distinct gesture proposals except for a slight similarity between deletion and replacement (having one gesture proposal in common).

For the **deletion** referents, 9 unique gestures were proposed in single and group referents with an AR of 0.08 for both. Three gestures appeared to be the most common among the participants. Those were: move *right index down* (Figure 1b), move *right index up*, and move the *right hand up* (Figure 1c). We decided to merge the index movement up and down into one gesture for two reasons: first, it is more intuitive to move the index finger up and then down (or down and up) because the user will have to move his hand back to a neutral position; second, participants p6 and p7 elaborated that moving the index finger up or down to delete is equally acceptable for them.

[2] https://rashad-j.github.io/conceptual-study

[3] Detailed results are shown on our website.

(a) Selecting a group of text items by distancing the index fingers.

(b) Deletion by moving the right index finger down.

(c) Deletion by moving the right hand up.

(d) Reordering by moving both hands simultaneously.

(e) Reordering by grabbing, moving, and releasing.

Figure 1: Common hand gestures for PE tasks proposed in our GES.

Moving the *right hand up* to delete was also common for the **replace** referent for both RP_s and RP_g. In general gestures for the **replacement** referent received a slightly higher AR of 0.10 and 0.18 for single and group referent respectively. Analyzing participants' thoughts, which were captured via think-aloud protocol, it appears that they wanted to delete first and then type the replacement item. Another common proposal for replacement was suggested by almost half of the participants (6/14), namely to simply type after selecting a text. Moreover, there were some proposals without agreement, e.g. p13 came up with the idea to *strike-through* text with the right index to delete and then type, whereas p14 suggested forming an "X" with his index fingers to delete before using the keyboard.

The **reordering** referents received three distinct gestures with AR of 0.16 and 0.26 for single and group referents respectively. The first one was to select and move the text with *both hands* by moving them simultaneously (Figure 1d). This gesture was proposed by 4 participants in RP_s and 6 participants in RP_g. The second gesture was to point with the *right index* finger and start moving it to move the text immediately after selecting (proposed by 4 participants in both RP_s and RP_g). The third gesture was to *grab* with the right hand and move the hand to reorder the text, then open it to release (Figure 1e). This gesture was proposed for RO_s by only 3 participants. Other individual proposals were made, e.g. p7 preferred to pinch using index finger and thumb, then move her hand to move the text, and then release the pinch to place the item.

Finally, the **insertion** referent received 5 unique gestures. One of the proposals was to point with the right index finger and then move it to place the cursor in the required place. This gesture was suggested by 9 out of 14 participants; hence, we see a high AR of 0.4. It was also referred to as *pointing* for single item selection. Once the cursor was placed in the target position, the user would switch to the keyboard for typing.

Together, these findings constitute a gesture set for text editing. Our separation into selection (for single items and groups) and editing operations makes the PE tasks more consistent and better represents our participants' mindsets. What is interesting is that selection of single items achieved high agreement on using a gesture to simply place the cursor on the item, without actually selecting it from start to end as with the mouse. The deletion and replacement referents shared some gesture proposals because participants often wanted to replace by deletion followed by typing. A further refinement to this set is presented below.

4 Prototype

We used the GES results to define our final gesture set and implement a prototype. For this, the frequently proposed gestures were explored in terms of implementation feasibility given the technology we are using. If two gestures were conflicting, we dropped the less popular one; otherwise we slightly modified it to resolve the conflict.

For **group selection**, we found that the proposed *index + thumb* gesture practically fails upon selection across multiple lines; thus, we dropped it. In contrast, using *both indices* can perform this kind of selection, so we implemented it as depicted in Figure 2. Note that in contrast to the mouse, the group selection using both index fingers allows the user to manipulate both ends of the selection continuously instead of having one side fixed. For **single item/position selection**, we only implemented *pointing* with the right index finger, as it already entails the *double tap* gesture. For multi-line text, both single and group selection allow pointing with the index finger vertically and horizontally.

Insertion can also be easily achieved by placing the cursor through *pointing* followed by typing.

For ***deletion***, D_s and D_g received similar gesture proposals. Looking at the proposals in detail, we found that two participants also wanted to delete

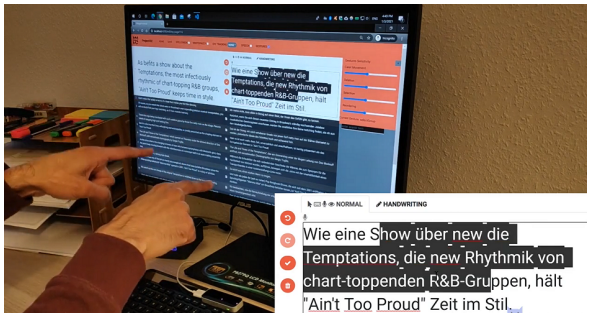


Figure 2: Mid-air gesture-based group selection by pointing with both indices.

with a hand down movement. Thus, we implemented *hand or finger movement down and up* to offer consistent deletion possibilities (Figure 3). For D_s, it is sufficient if the cursor is placed somewhere on the word; there is no need to define the start and end of the word through a group selection.

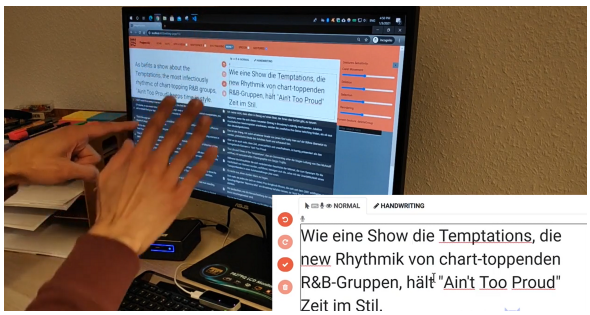


Figure 3: Mid-air gesture-based deletion by moving the right hand or finger up or down.

Replacement can be achieved by either performing a group selection and typing directly, or by selecting a single item or group of items, deleting, and then typing. Note that RP_s can thus also be achieved without group selection.

The most complicated gestures were proposed for ***reordering***; the gestures are a compound of several sub-gestures. Since reordering using the *right index* conflicts with cursor movement, we dropped it. Moving *both hands* while in the selection position turned out to be difficult to perform, as maintaining the same distance between the hands at all times is challenging. Therefore, we decided to merge it with the *grab* proposal; thus, after selection, a grab with the left hand indicates the start of the reordering process. Then moving both hands or just the right index finger reorders the text (Figure 4). Once the required position is reached, closing the right hand ends the reordering process and drops the text in the target position. For single item reordering, it is again sufficient to place the cursor on the item without selecting the whole text.

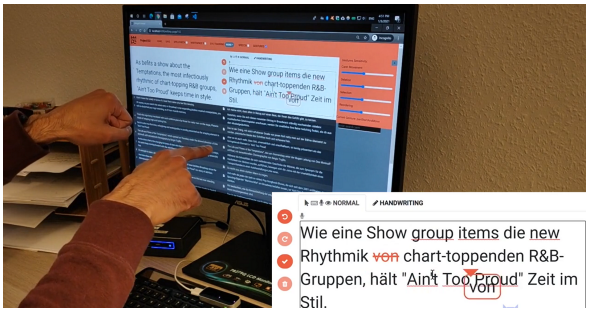


Figure 4: Mid-air gesture-based reordering by selecting, left grab, pointing with the right index finger to the target position, and releasing the grab.

The prototype was implemented as an extension to our open-source MMPE CAT interface (Herbig et al., 2020b,c)[4]. MMPE allows translators to use input modalities such as speech, touch, pen, and eye tracking in combination with the standard mouse and keyboard. However, it previously did not support mid-air gestures. The main interface shows the source on the left, and the target on the right, with the currently edited segment enlarged. This additional space turned out to be useful for hand gestures as it simplifies pointing. In addition, all user interactions are logged. MMPE uses Angular for the front-end, and node.js for the back-end, with WebSockets and REST APIs for the communication between them.

Our gesture detection relies on the *Leap Motion Controller*[5], which is small in size (8cm * 3cm) and can be placed on the top of the keyboard (Figure 2). The device provides frames of detected hands with 3D positions of finger joints, as well as some basic detection such as whether the fingers are extended or not. Based on this information our gesture detection algorithm determines if one of the above gestures is being performed. If only the right hand is detected with the index fingers extended, then the cursor will be updated based on hand movement. Moving both index fingers selects the corresponding text in the interface (Figure 2). When a deletion gesture is detected, the selected text (for group selection), or the word that the cursor is currently positioned on, is removed (Figure 3). A grab with the left hand puts the currently selected text/word

[4] https://github.com/NicoHerbig/MMPE
[5] https://www.ultraleap.com/product/leap-motion-controller/

containing the cursor in a reordering visualization. Then, movements of the right index are tracked and move the highlighted text as well as an arrow indicator visualizing the currently calculated drop position. Releasing the grab then places the text back into the input field at the indicated position (Figure 4). To avoid unintended gestures while moving the hands back to the keyboard, the user can form a grab in both hands after executing a gesture. Since people move their hands at different speeds, we further added sensitivity settings for gestures, similar to the standard mouse settings. A video showing the interactions in practice can be found under: https://youtu.be/qIRYeojkFVc.

5 Prototype Evaluation

In contrast to the web-based elicitation study, we had to evaluate the prototype in-situ due to the hardware setup. Given the COVID-19 situation, it was impossible to invite professional translators. Therefore we had to conduct a study with our colleagues.

To mitigate the difference between non-translation professional subjects (computer scientists) and translation professionals, we ensured that similar to professional translators, (i) all our participants have academic training (computing degrees instead of translation degrees), (ii) that they are also highly familiar with traditional mouse and keyboard interfaces and use them in their day-to-day work, (iii) all subjects have relevant language proficiency (source EN, target DE), and (iv) all work in a multilingual EN-DE environment. Furthermore, as the evaluation required participants only to perform pre-specified text editing operations, without involving any linguistic translation decisions, we hope to minimise the effect of not having translators as participants.

5.1 Method

We use a methodology similar to that of our previous MMPE evaluation (Herbig et al., 2020a), however, here we compare a novel interaction modality (mid-air hand gestures) to mouse and keyboard:

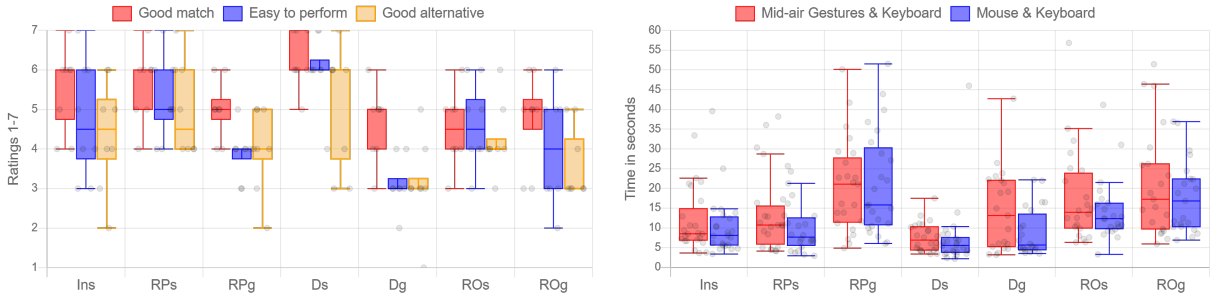
Participants: Overall, 8 participants (7 male, 1 female) from the department of computer science took part in the experiment: 5 researchers, 2 PhD students, and 1 MSc student. Their ages ranged from 24 to 39 (avg: 29, SD: 5). All had English skills from B2 to C1 and were either German natives (7 of them) or had C1 German knowledge. As computer scientists, they were all experienced keyboard users. Participants were all right-handed and had normal vision. Two of them indicated little experience with gesture-based interfaces, whereas the others reported a medium to very high level.

Apparatus: The main equipment consists of a 23 inch monitor, a NUC PC, a Leap Motion Controller, a standard wired mouse, and a standard keyboard with German layout. The NUC PC is equipped with a processor of type Intel(R), Core i7 CPU @ 3.50 GHz, 16.0 GB of RAM, and an internal graphics processor capable of capturing 30 – 60 frames per second when used by the Leap Motion Controller.

Procedure: Prior to undertaking the study, ethical clearance was obtained from the ethical review board at the university. The study consisted of 3 phases and took approximately 1 hour per participant. The first phase introduced GK and the prototype interface, followed by capturing demographic information. In the second phase, participants were given 10 – 15 minutes to explore GK to correct samples of incorrect MT output. The third phase included the main experiment, in which participants performed a guided test to correct MT output in two conditions: mid-air gestures & keyboard (GK) and standard mouse & keyboard (MK). For each of the referents from our elicitation study, 3 different segments had to be corrected in both conditions appearing in random order to capture comparable editing times. The segments were taken from the WMT EN-DE 2018 news test set. A single error was introduced per segment and a pop-up always told participants what error needed to be fixed and which modality to use. After each referent (e.g. deleting a single item), participants were presented with the same three 7-point Likert scales as in our GES. In addition we conducted semi-structured interviews to gather further feedback. We had 2 conditions, 7 referents, and 3 segments per referent; thus, there were in total $2 * 7 * 3 = 42$ segments to correct for each participant. While this correction of pre-defined errors prevents us from drawing conclusions in a realistic setting, it allows us to explore each editing operation in isolation, including accurate time measures and subjective feedback, which is more important for a first prototype test.

5.2 Results & Discussion

Qualitative data was collected by the semi-structured interviews and Likert rating scales after each referent. Figure 5a shows that operations manipulating single items were generally rated higher



(a) Subjective ratings.

(b) Editing duration of GK and MK.

Figure 5: Prototype evaluation results.

than operations on groups of items. D_s was rated best, especially in terms of goodness and ease of use. The majority of our participants commented that group selection was hard to perform, whereas the editing operations themselves were considered easy. While comments differed depending on the referent, most of them were positive, and we frequently got statements such as "it is great, [GK] felt like the same level of MK".

Quantitative data, shown in Figure 5b, captured the editing duration of both GK and MK for each referent, showing that the GK interquartile range was higher than the standard MK, except for RO_g. However, the most interesting finding was that, although the participants had years of experience using MK and were new to GK for text editing, the average editing time in the GK condition was very close to the average for MK in 4 out of 7 referents. For analyzing statistical differences in our data, we ran Wilcoxon signed-rank tests since the normality assumption of t-tests was not fulfilled due to the small sample size. As expected, given the limited amount of data, our statistical tests were unable to find significant differences between GK and MK for all operations[6].

Similar to what we found in the qualitative analysis, the gestures operating on single items were more efficient than operations on groups of items in the GK condition. D_s was the fastest, followed by RP_s and I. On the other hand, group operations turned out to be the most time-consuming in both conditions, with the biggest differences between conditions for D_g and RP_g. Interestingly, average editing time of RO_g was nearly identical in both conditions, although the gesture-based approach showed more variance.

[6] $\alpha = 0.05, \forall P, P > \alpha, P = (Ins = 0.641, RP_s = 0.312, RP_g = 0.945, D_s = 0.461, D_g = 0.461, R_s = 0.312, R_g = 0.383)$

In summary, the study has shown positive attitudes towards using mid-air hand gestures in combination with the keyboard for specific PE tasks. Single item referents in particular received good feedback and were close to MK in terms of time measures. Group selection was the main reason for disliking the GK and main source of additional editing time. Based on the comments, the majority of participants found such group selections difficult to perform, especially when selecting across multiple lines, therefore, improvements should be made to the group selection in the future. Overall, the results are encouraging, especially when considering the level of experience our participants had with MK and the short time for them to learn GK for text editing. In particular the single item referents, and perhaps improved versions of the group referents, could provide benefit to the PE process as a complement, not replacement, to traditional mouse- and keyboard-based editing.

6 Conclusion

The use of MT and PE changes the task of translation from mostly text production to fixing errors within useful but partly incorrect MT output. This affects the interface design of CAT tools, where translators need more support for text editing tasks. The literature suggests that other interaction modalities than MK, or combinations thereof, could better support PE operations. To the best of our knowledge, this is the first study that investigates the usefulness of mid-air hand gestures for PE of MT.

Our GES with 14 freelance translators yielded a set of gestures to manipulate both single items and groups of items, which we further refined by considering conflicting gestures and exploring them practically. The resulting prototype allows users to (i) place the cursor by pointing with the index finger, (ii) select ranges of text by pointing with both

index fingers, (iii) moving the hand or index finger up or down for deletion, and (iv) reorder by selecting text, forming a grab with the left hand, pointing with the right index finger to the desired position, and releasing the grab to drop the text. These gestures, combined with the keyboard, support all text manipulations required for PE.

Due to COVID-19, only a small-scale prototype evaluation with non-translator participants was possible. Nonetheless, as the prototype design was guided by an elicitation study with translation professionals which usually leads to well-perceived interfaces and since we designed the study to mitigate bias induced by a sub-optimal participant sample, we expect that professional translators would have given us comparable feedback. The findings overall suggest that GK could be a suitable interaction modality for PE and thus merits further research: Even though participants had years of experience with MK, our quantitative analysis of editing time showed that GK was only slightly slower for most operations, especially when manipulating single items. Similarly, qualitative data shows that manipulating single items was rated higher than operations working on groups of items, as participants found the group selection gesture "cumbersome" to perform. This finding indicates that further effort should be invested in improving group operations, which are also common in PE (e.g. by exploring if a different placement of the detection device could increase detection accuracy). However, the appealing results on single item operations and the satisfactory results on group operations bode well and warrant further exploration with professional translators in a realistic PE scenario.

We do expect that after using the interface for a longer period of time, users will become more effective, as is common with other interfaces: the new interface is competing with decades of MK muscle-memory training. However, only future long-term studies can show if editing times with GK will become as low as or even lower than with MK approaches. Apart from efficiency, participants in our previous studies (Herbig et al., 2019) argued for having multiple suitable options to interact with text, instead of performing the same movements all day long. Therefore, it is not just a question of speed but also user satisfaction and health: Additional modalities may help guard against carpal-tunnel syndrome and provide exercise alternatives in a seated environment.

To conclude, this new interaction modality, which so far was overlooked by research on CAT tools and post-editing, performs better than expected and therefore warrants further investigation. Overall, we hope that future research will pick up the insights from the first and second study and help advance the state-of-the-art in PE.

Acknowledgments

This research was funded in part by the German Research Foundation (DFG) under grant number GE 2819/2-1 (project MMPE). We thank all participants of the two studies for their valuable feedback.

References

Vicent Alabau, Christian Buck, Michael Carl, Francisco Casacuberta, Mercedes García-Martínez, Ulrich Germann, Jesús González-Rubio, Robin Hill, Philipp Koehn, Luis A Leiva, et al. 2014. CASMACAT: A computer-assisted translation workbench. In *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 25–28.

Susan Bassnett. 2013. *Translation*. Routledge.

Jan Van den Bergh, Eva Geurts, Donald Degraen, Mieke Haesen, Iulianna Van der Lek-Ciudin, Karin Coninx, et al. 2015. Recommendations for translation environments to improve translators' workflows. In *Proceedings of the 37th Conference Translating and the Computer*. Tradulex.

Michael Carl and Martin Kay Kristian TH Jensen. 2010. Long distance revisions in drafting and post-editing. *Natural Language Processing and its Applications*, page 193.

Sven Coppers, Jan Van den Bergh, Kris Luyten, Karin Coninx, Iulianna Van der Lek-Ciudin, Tom Vanallemeersch, and Vincent Vandeghinste. 2018. Intellingo: An intelligible translation environment. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13.

Spence Green, Jeffrey Heer, and Christopher D Manning. 2013. The efficacy of human post-editing for language translation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 439–448.

Nico Herbig, Tim Düwel, Santanu Pal, Kalliopi Meladaki, Mahsa Monshizadeh, Antonio Krüger, and Josef van Genabith. 2020a. MMPE: A multi-modal interface for post-editing machine translation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1691–1702. Association for Computational Linguistics.

Nico Herbig, Santanu Pal, Tim Düwel, Kalliopi Meladaki, Mahsa Monshizadeh, Vladislav Hnatovskiy, Antonio Krüger, and Josef van Genabith. 2020b. MMPE: A multi-modal interface using handwriting, touch reordering, and speech commands for post-editing machine translation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 327–334. Association for Computational Linguistics.

Nico Herbig, Santanu Pal, Tim Düwel, Raksha Shenoy, Antonio Krüger, and Josef van Genabith. 2020c. Improving the multi-modal post-editing (MMPE) CAT environment based on professional translators' feedback. In *Proceedings of 1st Workshop on Post-Editing in Modern-Day Translation at the AMTA Conference*, pages 93–108. Association for Machine Translation in the Americas.

Nico Herbig, Santanu Pal, Josef van Genabith, and Antonio Krüger. 2019. Multi-modal approaches for post-editing machine translation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–11.

Maarit Koponen. 2012. Comparing human perceptions of post-editing effort with post-editing operations. In *Proceedings of the Seventh Workshop on Statistical Machine Translation*, pages 181–190.

Maarit Koponen. 2016. Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *The Journal of Specialised Translation*, 25:131–148.

Kaisa Koskinen and Minna Ruokonen. 2017. Love letters or hate mail? Translators' technology acceptance in the light of their emotional narratives. In *Human Issues in Translation Technology*, pages 26–42. Routledge.

Panayiotis Koutsabasis and Panagiotis Vogiatzidakis. 2019. Empirical research in mid-air interaction: A systematic review. *International Journal of Human–Computer Interaction*, 35(18):1747–1768.

Elina Lagoudaki. 2009. Translation editing environments. In *MT Summit XII: Workshop on Beyond Translation Memories*.

Samuel Läubli, Sheila Castilho, Graham Neubig, Rico Sennrich, Qinlan Shen, and Antonio Toral. 2020. A set of recommendations for assessing human–machine parity in language translation. *Journal of Artificial Intelligence Research*, 67:653–672.

Joss Moorkens and Sharon O'Brien. 2017. Assessing user interface needs of post-editors of machine translation. In *Human Issues in Translation Technology*, pages 127–148. Routledge.

Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, MC Schraefel, and Jacob O Wobbrock. 2014. Reducing legacy bias in gesture elicitation studies. *interactions*, 21(3):40–45.

Michael Nielsen, Moritz Störring, Thomas B Moeslund, and Erik Granum. 2003. A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In *International Gesture Workshop*, pages 409–420. Springer.

Sharon O'Brien. 2012. Translation as human–computer interaction. *Translation Spaces*, 1(1):101–122.

Michael Ortega and Laurence Nigay. 2009. Airmouse: Finger gesture for 2d and 3d interaction. In *IFIP Conference on Human-Computer Interaction*, pages 214–227. Springer.

Sharon O'Brien, Joss Moorkens, and Joris Vreeke. 2014. Kanjingo–a mobile app for post-editing. In *Proceedings of the 17th Annual Conference of the European Association for Machine Translation (EAMT 2014), Dubrovnik, Croatia, 16th-18th June*.

Mirko Plitt and François Masselot. 2010. A productivity test of statistical machine translation post-editing in a typical localisation context. *The Prague Bulletin of Mathematical Linguistics*, 93(1):7–16.

Maja Popovic, Arle Lommel, Aljoscha Burchardt, Eleftherios Avramidis, and Hans Uszkoreit. 2014. Relations between different types of post-editing operations, cognitive effort and temporal effort. In *Proceedings of the 17th Annual Conference of the European Association for Machine Translation*, pages 191–198. European Association for Machine Translation Dubrovnik, Croatia.

Christopher M Rives, Craig T Brown, Dustin L Hoffman, and Peter M On. 2014. Gesture based edit mode. US Patent 8,707,170.

Caroline Rossi and Jean-Pierre Chevrot. 2019. Uses and perceptions of machine translation at the european commission. *The Journal of Specialised Translation (JoSTrans)*.

Ram Pratap Sharma and Gyanendra K Verma. 2015. Human computer interaction using hand gesture. *Procedia Computer Science*, 54:721–727.

Carlos S.C. Teixeira, Joss Moorkens, Daniel Turner, Joris Vreeke, and Andy Way. 2019. Creating a multimodal translation tool and testing machine translation integration using touch and voice. *Informatics*, 6.

Antonio Toral, Martijn Wieling, and Andy Way. 2018. Post-editing effort of a novel with statistical and neural machine translation. *Frontiers in Digital Humanities*, 5:9.

Olga Torres-Hostench, Joss Moorkens, Sharon O'Brien, and Joris Vreeke. 2017. Testing interaction with a mobile mt post-editing app. *Translation & Interpreting*, 9(2):138–150.

Radu-Daniel Vatavu and Jacob O Wobbrock. 2015. Formalizing agreement analysis for Elicitation Studies: New measures, significance test, and toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1325–1334.

Mihaela Vela, Santanu Pal, Marcos Zampieri, Sudip Kumar Naskar, and Josef van Genabith. 2019. Improving CAT tools in the translation workflow: New approaches and evaluation. In *Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks*, pages 8–15.

Juan Pablo Wachs, Mathias Kölsch, Helman Stern, and Yael Edan. 2011. Vision-based hand-gesture applications. *Communications of the ACM*, 54(2):60–71.

Frank Weichert, Daniel Bachmann, Bartholomäus Rudak, and Denis Fisseler. 2013. Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, 13(5):6380–6393.

Jacob O Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A Myers. 2005. Maximizing the guessability of symbolic input. In *CHI'05 extended abstracts on Human Factors in Computing Systems*, pages 1869–1872.

Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1083–1092.

Julián Zapata. 2016. Translating on the go? Investigating the potential of multimodal mobile devices for interactive translation dictation. *Revista Tradumàtica: tecnologies de la traducció*, 14:66–74.