Investigating the Effect of Emotive Translation Bubbles During K-Pop Live Streams

Final Report

ZAYN JAMEEL ABBAS, University of Guelph, Canada
PHAM SKY TRUONG, University of Guelph, Canada

1 INTRODUCTION

Over the years, automatic translation has gained widespread use. Automatic translation has several well-known applications, from academic to commercial. While integrating machine translation is beneficial, they only sometimes produce optimal results. This paper investigates how improving machine translation output through interface design can boost content comprehension. We focused on the K-pop fandom, as these fans are found globally and speak various languages. The K-pop fandom offers an inclusive community, allowing fans with different identities, abilities, and cultures to connect and their idols. In addition, K-pop idols frequently use social media live streams to interact with their followers. Unfortunately, many live streams are conducted in Korean, with translations following days to a week after the initial stream, making it difficult for viewers who need help understanding Korean to participate.

Previous research has explored methods to enhance machine translation's perceived clarity and overall quality. Speech bubbles added to text-based chat messages have been found to aid in clarifying emotional meaning and improving content understanding outside of the context of translation. We performed an informal study where we mocked up emotive translation bubbles, which represent the emotions expressed by the idol, to analyze the impact these bubbles had on the user's live stream experience.

1.1 Objectives

Our ultimate objective was to determine if such a design feature could improve user experience with machine-translated text when used for live stream scenarios. To accomplish this, we want to know how consumers who do not read, write, or speak the Korean language approach K-pop content. We wanted to find out the expectations and the levels of expectation the consumers have towards using a feature designed to increase their experience.

1.2 Scope

Our four-month research project on Investigating Multilingual Language Support for Live Streaming Popular Culture Media Content taught us that developing a machine learning model to improve translation accuracy is complex. It is a lengthy process which demands time and energy. Some challenges we discovered during our research include capturing audio, speech recognition, audio transcription and text-to-text translation.

We wanted to expand our understanding of the machine translation world for this research term. However, we wanted to uncover what the users desired instead of chasing a highly accurate model. This research term included discovering methods to improve machine translation output, designing user interfaces that included one or more of these methods, and conducting a pilot study followed by an informal user study. Due to the limited timeframe, we

decided to move forward with one final design for the user study, the emotive translation bubbles.

1.3 Research Plan

The first step was to explore different supplementary techniques that can be used to improve machine translation output; some of these included speech bubbles, emoticons, and highlighting keywords [2]. We then picked our top four techniques to design a mock user interface of each and a combination of them. This resulted in a total of seven low-level designs. Figma was used to create a high-fidelity design of our top three designs in our next step. Once we decided on a final design, we designed an informal pilot user study. Lastly, we conducted four user studies using participants in the K-pop fandom and analyzed their results.

We decided to continue our translation of choice to be Korean to English. For this research, it meant that we would be sourcing for recorded live streams of one or multiple K-pop idols speaking in Korean where English translations had been made available. K-pop idols are referred to as music artists. This identified our candidate pool as those interested in K-pop culture or who have some familiarity with Korean media and who could read and speak in English with moderate to good influence.

Although K-pop idols mostly stream on a platform called Weverse® (previously V-Live®), we found that YouTube® had an excellent collection of post-live videos and utilized both platforms to search for content. In the scope of this research, the eligible content is limited to solo streaming appearances to avoid added layers of complexity from pair or group streams. The video content had to be the K-pop idol (streamer) interacting with fans and possibly doing other activities. We refrained from choosing edited videos to avoid causing confusion with our intended design. This resulted in selecting videos where the English translation could be turned on/off rather than being overlayed.

As the research this term focused on the design aspect rather than programming, we did not place constraints on the technical side. For instance, during brainstorming sessions, we would like to generate as many concepts for supplementary techniques as possible and not worry about if there exists a technology to implement it.

2 BACKGROUND RESEARCH

2.1 Previous Study

Within our previous study, we implemented text-to-text translation by feeding our dataset, consistent with Korean children's stories and song lyrics, to a No Language Left Behind (NLLB) pre-trained model and produced reasonable results. A challenge we ran into was the repetition of words for specific data inputs, mostly for song lyrics. We created a script to translate the original Korean text line by line to overcome this issue. This technique improved the accuracy of the translations and brought our awareness of the importance of punctuation.

To evaluate our translations, we implemented the BLEU scoring metrics. The scores ranged from 10%-50% accuracy; the translations could provide the readers with the gist of the content. However, more future work needs to be done to improve the accuracy of the

translations. The foreseeable milestones stemming from the work of this research include investigating the cause of repetition and the effect of punctuating a sentence or sentence.

Researching and working with the NLLB model have given us more profound insights into the challenges of developing an NLP model and learning to develop creative solutions to tackle those.

2.2 Related Work

Many techniques can be used to improve the output of machine translation. We looked into eleven of these techniques to determine which would be best suited for live streams. *Table 1 (Techniques for Improving Translation Output)* summarizes all eleven techniques and their pros and cons. Based on these pros and cons and the constraints and costs of grounding, we determined that our top four choices to move forward with were highlighting keywords, clues about translation quality, annotations via emojis, and predicting intentions via speech bubbles.

Technique	**Pros**	**Cons**
Annotations with cultural and emotional information [2,4,5]	Assists with understanding cultural context of translations.	Can create a cluttered user-interface in a live setting.
Back translation [2,7,8]	Improves the quality of the translations.	Not applicable/beneficial for the purposes of live commercial translation.
Highlighting keywords [2,9,10,11]	Provides users a point of focus.	Could become distracting. The words highlighted are subjective from user to user.
Post editing [2,12]	Improves the accuracy and quality of translations.	Not applicable in real-time translation.
Clues about translation quality [2,13]	Provides users with a confidence about the quality of the translation.	Possibility of causing confusion and frustration as the user interface may become too cluttered.
Showing two outputs [2,6]	Increases the confidence in the translation's accuracy.	Can become overwhelming for users on the receiving end.
Automatic Evaluation of Subtitles	Beneficial for researchers to assess the translation accuracy.	Not applicable/beneficial for the purposes of live commercial translation.
Annotations via emojis	May help to improve the understanding of the emotions behind the context of the translations.	There is a large variety of emojis. People may use the same one to express multiple emotions.
Predict intentions via speech bubbles	May help to improve the understanding of the emotions behind the context of the translations.	Must predetermine which speech bubbles depict which emotion.
Closed captions [2,13]	Improves the accessibility of live videos.	Could become overwhelming for some viewers.

Table 1: Techniques for Improving Translation Output

The highlighting keywords technique involves identifying and highlighting important words or phrases in the source text to improve the accuracy of the translation output. Gao et al. (2013) found that highlighting keywords in the source text improved the quality of the translations and significantly impacted the user experience of the translation process. Participants in their study preferred highlighted translations and felt more confident in the

accuracy of the translations when keywords were highlighted.

Similarly, other studies found that by identifying and prioritizing important words and phrases in the source's text, machine translation systems can better capture the intended meaning of the original text and produce more accurate and fluent translations [9,10,11].

The clues about translation quality technique involve providing users with contextual clues about the quality and reliability of the translation. Hara and Iqbal (2015) conducted a formative study to investigate the effect of machine translation on interlingual conversations. They found that participants were more likely to trust and use the machine translation output when provided with contextual clues about the quality of the translation. These clues included information about the translation's accuracy, the translation system's confidence level, and the potential sources of error in the translation process.

Providing contextual clues about the translation quality is a promising approach for improving machine translation systems' user experience and effectiveness [2, 13]. These clues can include information about the source language, the translation system used, and the potential errors or limitations of the translation output [2, 13]. This technique can help users better understand and trust the machine translation output by providing contextual information about the translation quality, leading to more effective and efficient communication in interlingual settings.

Studies have evaluated the use of emojis to provide feedback on the quality of machine translation output in a human-in-the-loop scenario [14]. They found that emojis can effectively capture user perceptions of translation quality and provide useful feedback for improving the accuracy and fluency of the translations. It could be beneficial to flip this around and use the emojis to reflect the emotion behind the context of the translation. Although we did not continue this path, it would be an interesting direction to explore in future work.

Paralinguistic cues refer to the nonverbal aspects of communication that accompany spoken or written language, such as intonation, facial expressions, and body language [15]. In a non-translation context, adding paralinguistic cues in the form of expressively shaped speech bubbles to text-based chat messages has enhanced communication by providing visual cues to support the understanding of the intended emotional tone and meaning of the message [15]. The use of paralinguistic cues can be especially important in contexts where communication is limited to text-based messages and lacks the nonverbal cues present in face-to-face communication. By adding speech bubbles to text-based chat messages, the sender can more effectively convey the intended emotional tone and meaning of the message, leading to improved understanding and a more engaging communication experience for both the sender and the recipient.

This approach could be adapted and expanded to live translation models by incorporating the use of speech bubbles in real-time translation systems. Speech bubbles could be used to represent the speaker's intended emotional tone, such as indicating sarcasm, humour, or emphasis. This would help to convey the intended emotional meaning of the message and improve the accuracy of the translation. Moreover, speech bubbles could be used to represent visual cues such as indicating whether the speaker is asking a question, making a statement, or expressing uncertainty. Live translation models could provide users with more intuitive and informative feedback on the conversation context by predicting the speaker's intention and representing it visually.

3 METHODOLOGY

3.1 Experimental Designs

The design went through many stages from tablet sketch to whiteboarding to Figma iterations. The very first sketches were made to align the general idea of where the main components would be.

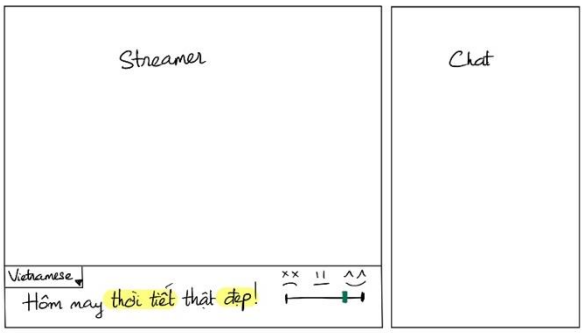


Figure 1. A: Sketch 1

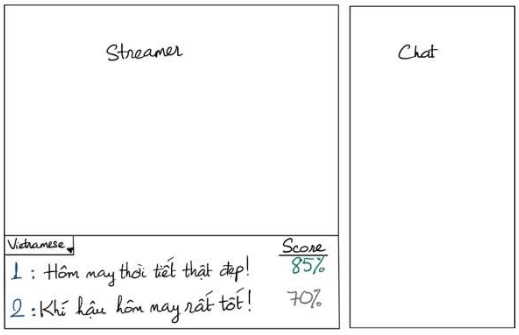


Figure 1. B: Sketch 2

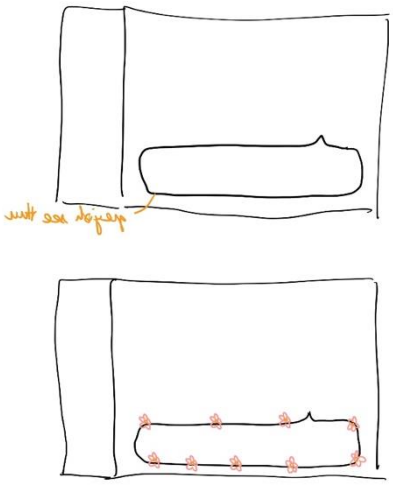


Figure 1. C: Sketch 3

The placement of streamer and chat views were designed after the live stream formats of YouTube® and Twitch®. At this point, the supplementary technique was implemented at the bottom of the screen, below the streamer, where captions and translated texts are commonly seen throughout various platforms. Sketches illustrated in Figures 1a and 1b depicted ideas of hybridization between different supplementary techniques evaluated from Section 2.2 of this research. Sketch 1 illustrates a synthesis between *Highlighting keywords* and *Annotations via emojis* while Sketch 2 shows a meshed-up version of *Showing two outputs* and *Automatic Evaluation of Subtitles* which was a percentage score. Sketches in Figure 1c showed initial design concepts that implemented *Predicting intention via speech bubbles* technique.

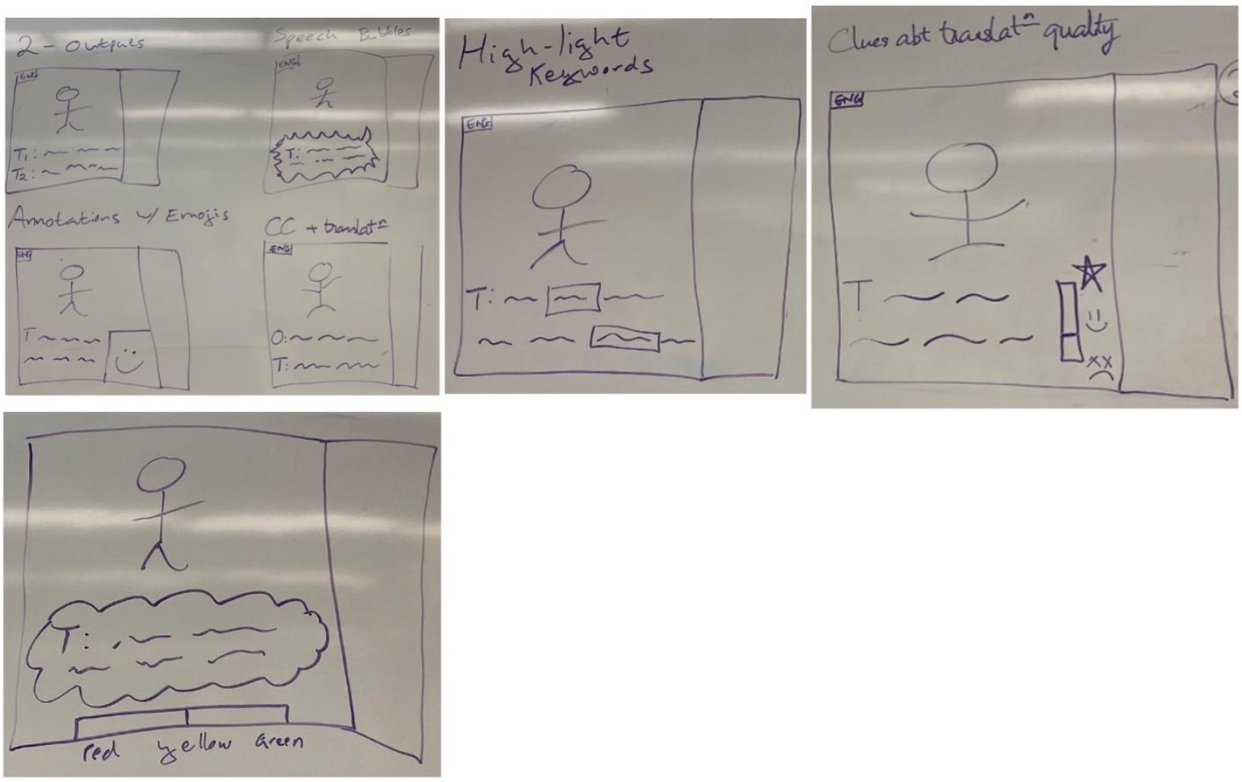


Figure 2: Whiteboard Sketch

With the layout of the main components agreed on, we brainstormed designs for each technique or hybridization of two or more techniques. The whiteboard sketches propelled us further into design consideration in addition to the techniques' evaluation discussed in Section 2.2. One of the criteria was to avoid a cluttered design with too much information. For example, *showing two outputs* and the combined technique of *Closed Captions* and showing translation would show too many texts in live stream settings. We predicted that the users would not have enough time to read all the different texts for the techniques to be helpful. We opted for designs that had more visual cues. Three designs that stood out and were selected for Figma iterations were *Highlighting keywords*, *Clues about translation quality* and *Predicting intention via speech bubbles*.

We mocked up ten panels in Figma. The panels were in pairs for designs showing and hiding the Chat component. The first pair was our Base template. The second and third pairs illustrated the Speech Bubble and Highlight keywords UIs respectively. The fourth and final pairs show different interpretations of *Clues about translation quality* design, with one using a colour scale and the other using an emojis scale. Figure 3, on the following page, shows design pairs side by side.

The Figma design gave us solid insights into the users might be experiencing. We quickly concluded that both of our design interpretations for *Clues about translation quality* would not fare well in a live stream setting despite being visual cues. Assuming we had an excellent ML model that could intelligently parse sentences based on speech, there were still issues with deciding how many sentences should be grouped together to calculate a score. Depending on the speech's speed and sentence length, the scale could constantly move without allowing users to comprehend the translation and the score. Furthermore, a low score might confuse the translation instead of offering meaningful assistance.

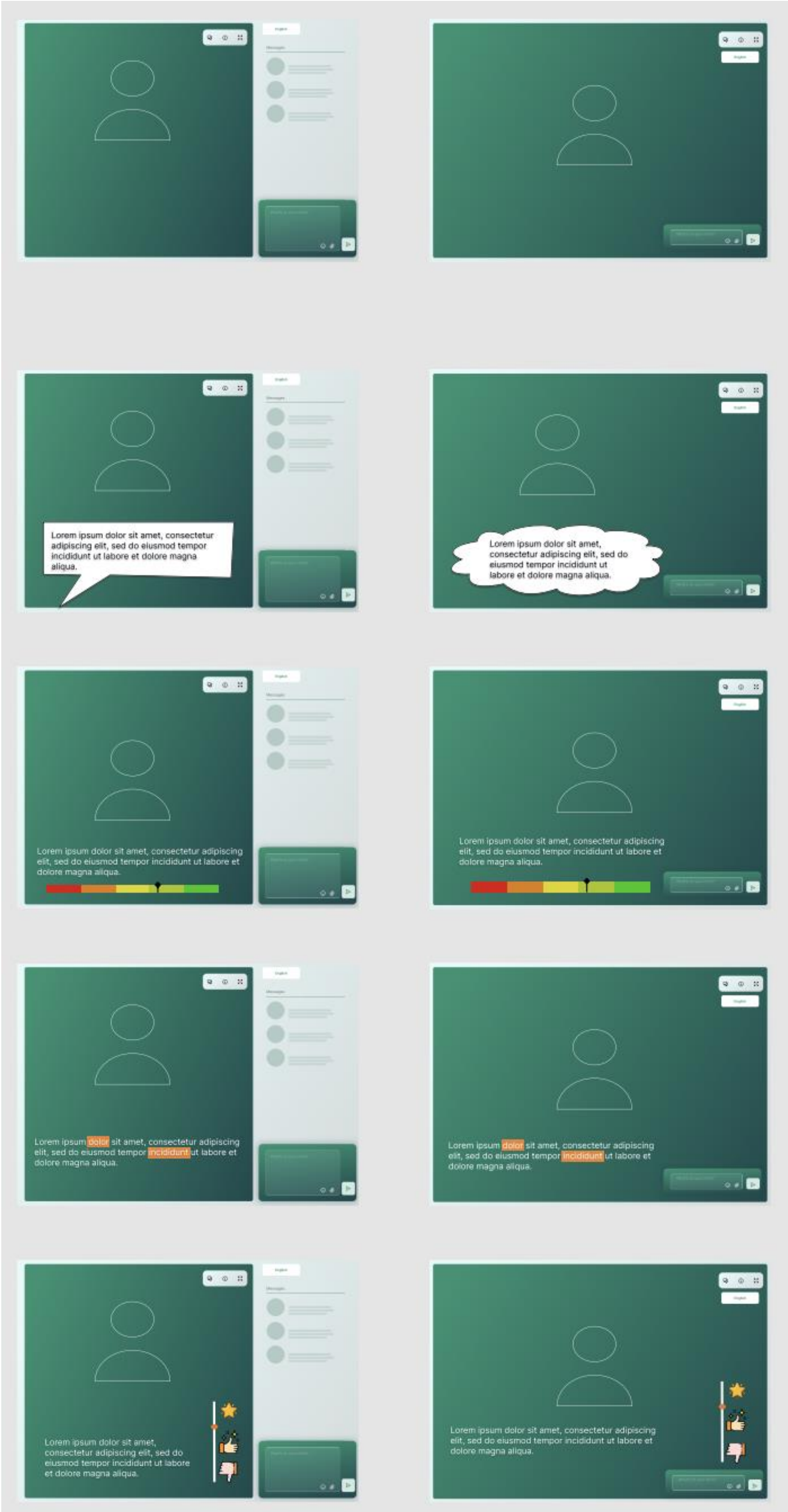


Figure 3: Figma Design

We were left with the Speech Bubbles and Highlight Keywords as potential candidates for full prototype design. These supplementary techniques were well integrated into the translated text. There were many possible design iterations for both techniques given the task of sourcing for existing speech bubbles or designing our own while deciding the intentions for highlighted keywords and how different highlight colours could lead to different user interpretations. Factoring in time constraints, we designed a prototype for only one supplementary technique. At this point, we decided to concentrate on conveying the emotions of a streamer's translated speech-to-text. We were aware that comics use speech bubbles to express a character's action for yelling, whispering, and thinking amongst other representations. Coupled with our perspective on how the speech bubbles design presented a clear visual despite seeming to have a seamless integration with the inner text, we picked Speech Bubbles as the only candidate to get a design prototype.

While looking for existing speech bubble designs for our project, we needed more than the variation we came across for our purpose. Since speech bubbles are traditionally used in comics, the speech bubble shape usually describes a character's state or action rather than different emotions. For instance, regular conversations or normal speech is usually depicted by oval, rounded-rectangle, or sharp-rectangle shapes with a tail; text inside a cloud-shaped bubble points to a character's inner thoughts; and a spiky-shaped bubble is used when a character is yelling [16, 17, 18]. We found a source that used an inverted speech bubble, a black bubble with white text, to show negative emotions [16]. These findings were reasonable since comic artists would draw characters expressing their emotions rather than relying on speech bubbles. We concluded it was necessary to assign our own definitions to some existing speech bubbles and design others to fulfill our needs. Figure 4 shows our custom speech bubbles and their definition assignment.

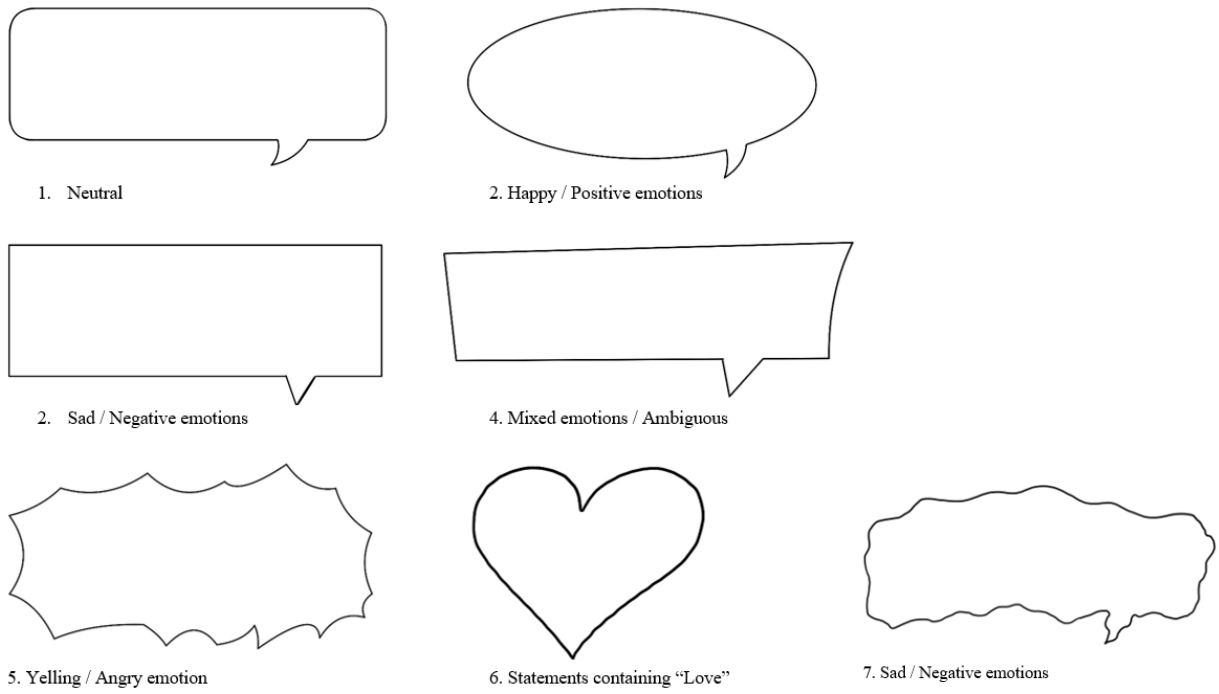


Figure 4. Speech Bubble Design

We had two iterations, speech bubbles two and seven, showing sad and negative emotions.

Speech bubble two was the first iteration where we tried to enforce a trend between shape and emotion. The sharper shape would depict a negative feeling; hence, a sharp-corner rectangle. The positivity increases as the shape is rounded out for rounded-rectangle and oval shapes. However, the changes were deemed subtle when we incorporated the speech bubbles with translation text and put them onto the content panels. Speech bubble seven's shape was modified to have a droopiness form, and the general design was inspired by speech bubbles that depict weakness [18].

With the purpose of implementing emotive translation bubbles, we searched for live streams where the K-pop idol expressed diverse emotions. Within the allotted time of content searching, we could not find a live stream where speech bubble five (Yelling / Angry emotion) could be utilized. We were convinced that at the stage of this research, the other emotive translation bubbles were sufficient to perform a user study. Two live streams that made it to the final of our list were from K-pop idols Suga (BTS) and Ryujin (Itzy) [19, 20].

To generate content, we went through each video and took a screenshot of all frames that have different text and different movements or facial expressions from the streamer and put them in chronological order. We had around 120 panels and laid them over the template (Figure 5).

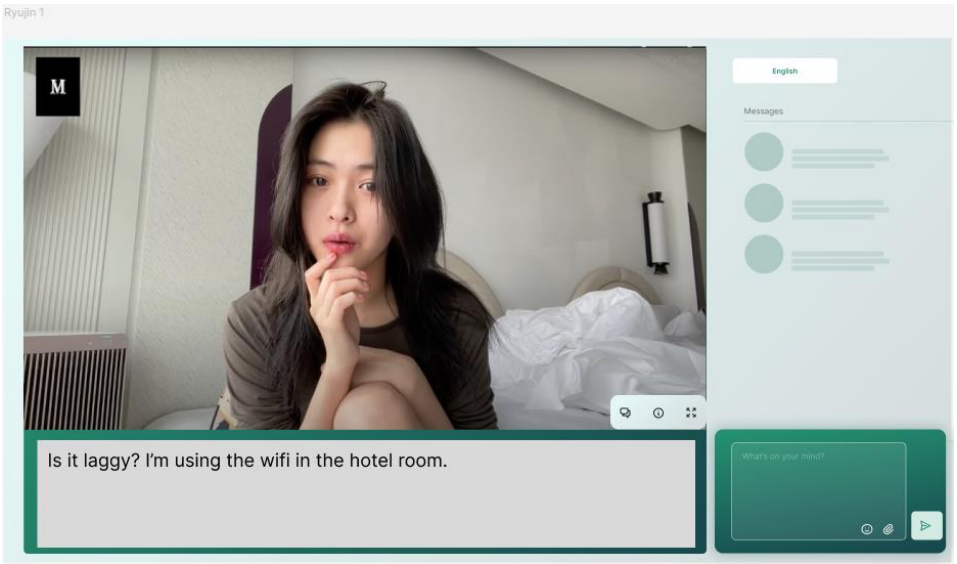


Figure 5. Early Design

Rather than showing the participants full live streams overlaid with speech bubbles, we wanted to pause in between and pose the questions from our user study to the participants. Letting participants view the streams in smaller portions would allow us to get their perspectives when the experience was fresh. We identified "stories" within the stream by grouping the panels where the streamer chatted with their viewers on a particular topic. From what we gathered going through the live streams, a story stemmed from the streamer answering a question by the viewer.

We selected three stories from each streamer for a total of six stories. From the male streamer, Suga, we identified three stories with topics on a shoulder injury, noodles, and coffee. From the female streamer, Ryujin, we identified three topics comics, movies, and travel. We assessed that the streamers expressed many emotions speaking on these topics and displayed diverse body movements and facial expressions.

In the spirit of the experiment, many of the panels were cut out to make the stories flow

smoothly. In addition, we wanted to increase the variety of speech bubbles the participants could experience at a faster pace. For example, a number of panels were merged together where their sentences contained a similar emotion which happened to be mostly neutral. Other panels were combined into one where the streamer's facial expression or body moment was consistent. In some cases where Suga chatted about noodles at two different points in the video, we combined the panels to make one coherent story. After the trimming activity, we were left with 64 panels.

When it was time to overlay the bubbles, we moved them to the top, above the streamer. We noticed that speech bubbles traditionally have a tail that points to the speaking character, specifically the mouth. To imitate that, we brought bubbles on top which were closer to the streamers' faces. Figure 6 shows the design we used for our Pilot Study. The background colour for the speech bubble had been changed to a shade of grey. The full content we used for the Pilot Study is in the attached zip file Design_Content.zip.

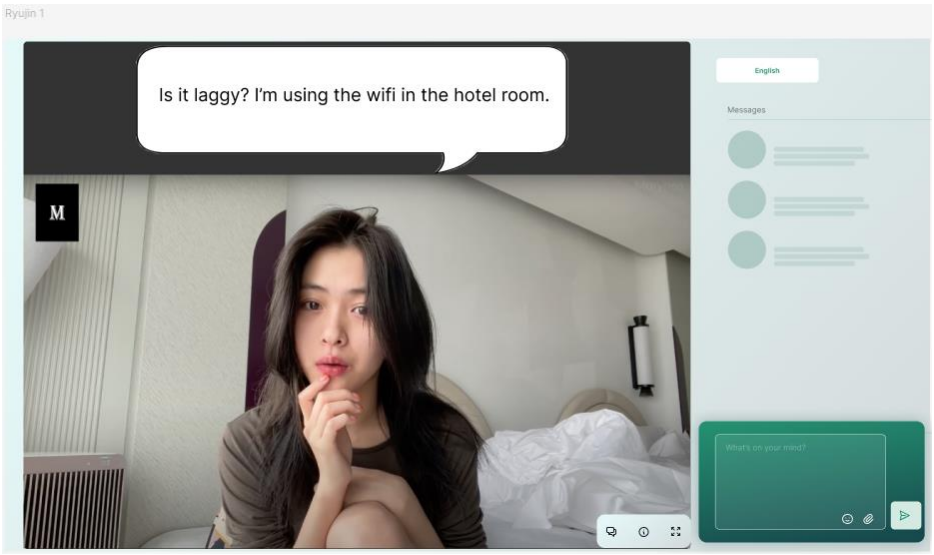


Figure 6. Pilot Study Panel

We used Figma's available features to link the panels together to create an interactive user study. Within a story, the panels were timed to switch based on the visual text length inside the speech bubble. The general rule of thumb was 6 seconds for 3 newlines, 4.5-5 seconds for 2 long newlines, 4 seconds for 1 long sentence and 3 seconds for shorter sentences. A sentence was considered long if its length was the length of the speech bubble, and a short sentence was half its length. Since we paused for questions and answers (Q&A) in between the stories, the last panel of each story persisted until being clicked on, then the new story started.

Until this point, we had estimated a fitting length for the study session to be under 45 minutes. We stayed mindful of the design steps of content generation to filter the panels to be concise and meaningful. To evaluate our study design, we ran a Pilot Study which is discussed in Section 3.2.2.

Before the Pilot Study, we were in the process of creating two study sets, Set 1 and Set 2, as a means of randomization. The study sets in PDF format can be found in the attached Design_Content.zip file. Set 1 and Set 2 have the Control story as the second story for each streamer. The Control story used the text container from the early design phase, as shown in Figure 5. Unfortunately, we did not have the set ready for the Pilot Study; hence the reason the Pilot Study had its own set.

After the Pilot Study, we created two more study sets, Set 3 and Set 4. Set 3 and 4 have the

Control story as the first story for each streamer. The design for all study sets was finalized at this point. Figure 7 shows the first panel of the first story for streamer Ryujin in the control format. Figure 8 shows the same using an emotive translation bubble.

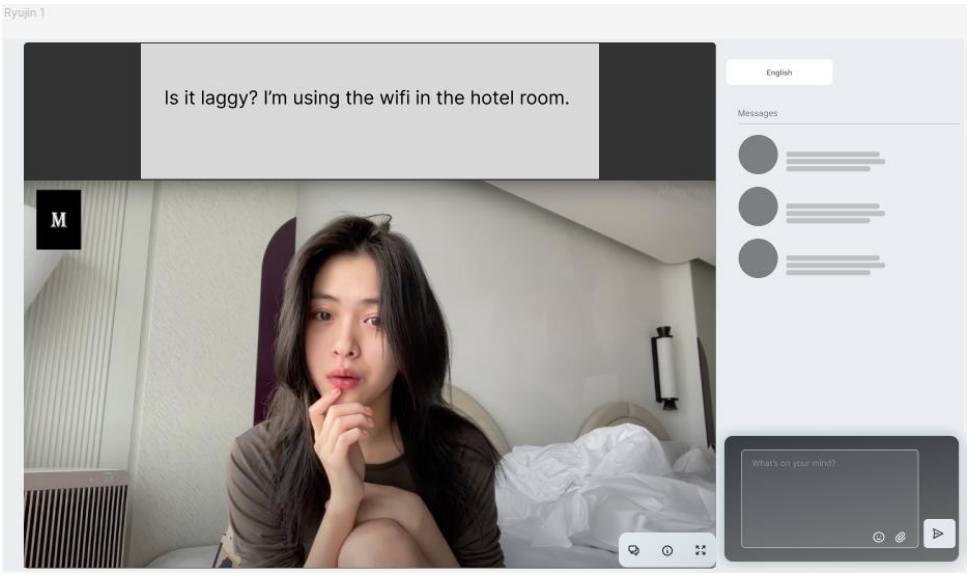


Figure 7. Control Panel

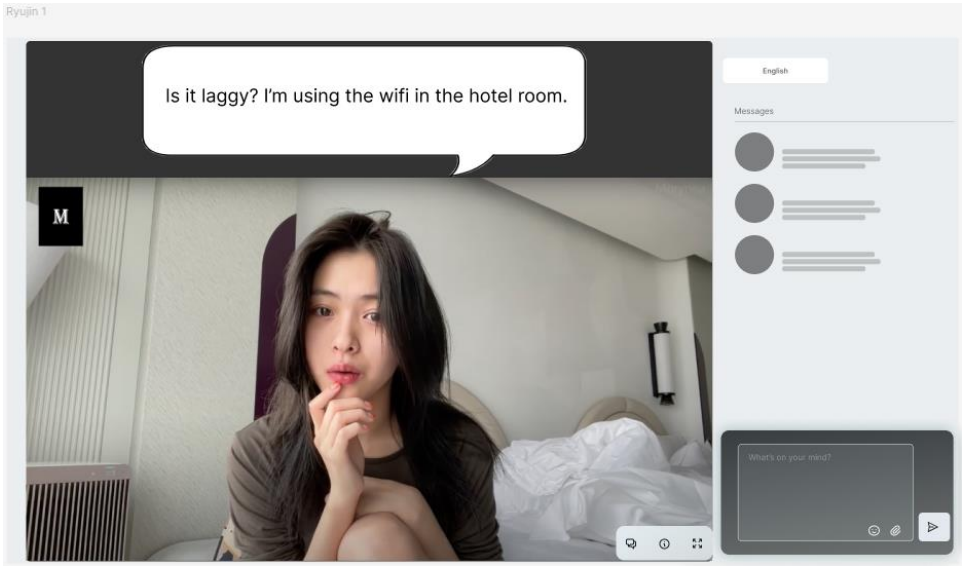


Figure 8. Speech Bubbles Panel

3.2 User Study

We decided to conduct an informal user study to determine the impact of emotive translation bubbles on the user's experience when watching live streams. Our research question is as follows:

"Does incorporating the emotive translation bubbles user interface for live K-Pop streams improve the understanding and experience between the idol and user?"

Some associated questions that could be followed up on include:

- Does the use of an **emotive translation bubbles** user interface during live K-Pop streams result in higher levels of engagement and interaction between the idol and the user, as measured by comments, likes, or shares?
- Are users who are given an **emotive translation bubbles** user interface during live K-Pop streams more likely to feel connected to the idol and the performance compared to those who aren't given the interface?
- Does incorporating an **emotive translation bubbles** user interface during live K-Pop streams lead to higher levels of satisfaction and enjoyment among users?
- Are there any significant differences in the understanding and experience of live K-Pop streams among users of different ages, genders, or nationalities when an **emotive translation bubbles** user interface is incorporated?

A complete document of our study can be found in *Appendix A.*

3.2.1 Planning the Study

One of the first stages in planning our study was to determine what the variables would be. The dependent variable is the user's experience. This included levels of engagement and understanding. The independent variable is the user interface for translation (the emotive translation bubbles). We then generated an ideal proto-persona. This proto-persona had no knowledge of Korean but was interested in the K-pop community and watched K-pop streams occasionally, where they relied heavily on translation to gain an understanding. Their goals included developing a better understanding of the topics of conversation in their community and increasing their connection to their idol that has been restricted due to language barriers.

Due to time constraints, we had to recruit participants that we knew. However, to ensure accurate results, our participants must fit the proto-persona. We decided on these questions:

1. Q: Do you watch idols live stream?
2. Q: Do you interact with idols on the live stream?
3. Q: Do you use translation applications/models to further immerse yourself in the K-Pop culture?

These questions helped us ensure that our participants aligned with our proto-persona. Our first question ensures that the participant has some knowledge of the K-Pop industry to an extent. The second question ensures that the participant may have interest in participating in live streams with their idols, despite language barriers. The last question indicates whether the participant will be an ideal candidate to use translation models to immerse themselves within the community.

Other preparatory setup steps included creating Microsoft Teams meeting links and structured documents for note-taking and observations. We generated Figma designs of our interface to walk participants through. We also generated a script to ensure that the study ran as smoothly as possible. This script can be found in *Appendix A*. The study itself was planned to have the user watch two individual streamers. Each of these streamers had three stories or topics of conversation. For each streamer, we had a control case for one of the stories and the remaining two were our emotive translation bubbles. In between each story, we planned to ask the following questions:

Q1: What were the emotions expressed in this story?
Q2: Does the translation widget accurately match the content?
Q3: Do you have any other comments?

After each streamer, we planned to ask the following:

Q1: How much did the speech bubbles impact your understanding on a scale of 0 to 5? (0 being not helpful, 5 being extremely helpful.)
Q2: Were you able to attend to the translation as well as the video scene?
Q3: Are there any other comments you have on this stream?

Lastly, at the end of the study, we set up eight questions:
Q1: Can you describe how the use of a speech bubble user interface during live K-Pop streams may affect your experience during the live stream?
Q2: Can you describe how the use of a speech bubble user interface during live K-Pop streams may affect your understanding of the stream?
Q3: Could you describe whether the use of speech bubbles would impact your sense of connection to the idol?
Q4: What current strategies do you use to understand live streaming content?
Q5: Display each speech bubble and ask them to comment on what they think they represent.

The last three questions were planned to be asked after we displayed the design of our emotive translation bubbles to the users.
Q6: Do you have any feedback about the design of speech bubbles?
Q7: Do you have any feedback on the study session?
Q8: Do you have any other comments?

For the study, the plan was to have Abbas carry out the introduction and facilitate the questions and answers (Q&A) session while Truong would act as the note taker. The Figma study set would be viewable through Truong's or the participant's screen sharing. If the participant was to share the screen with us, we would provide the link to the Figma study set and give instructions to start or continue a story during the study. If Truong was to share the screen, then the researchers would coordinate to give each other prompts to start the study, a story or to announce the end of a story or a stream.

To ensure all preparations were complete, we ran a pilot study with Leila Homaeian, a Ph.D. in Systems Design Engineering at the University of Waterloo. The pilot study was scheduled a week before other user studies, so any last-minute changes and considerations would be accounted for.

3.2.2 Pilot Study

Our pilot study with Homaeian indicated that the user study was feasible but required adjustments. We could confirm that the length of one study session was well within 45 minutes. The pilot study lasted over 30 minutes with a complete user study, clarification questions and suggestions from Homaeian. The pilot study made us realize some panels were not linked properly. The incident made it clear that the screen-sharing should be done by one of the researchers to react promptly during any technical issue. This would ensure the participant focuses entirely on the speech bubbles experience or lack thereof.

One insightful recommendation was displaying a grid view of the entire streamer story for each story Q&A. Homaeian pointed out that this would be beneficial as the user could have a reference while answering our questions. Figure 9 shows a grid view of story 2 from Suga. A full list of grid views for each story can be found in the Design_Content.zip file. Another

suggestion was to label our emotive translation bubbles by number so it would be easier to refer to each individual bubble during Q&A. We also discussed the necessity of randomization amongst the study sessions and emphasized the need to create study sets with the Control as starters. In addition, we polished our questions to be clear, concise and bias-free as much as possible.

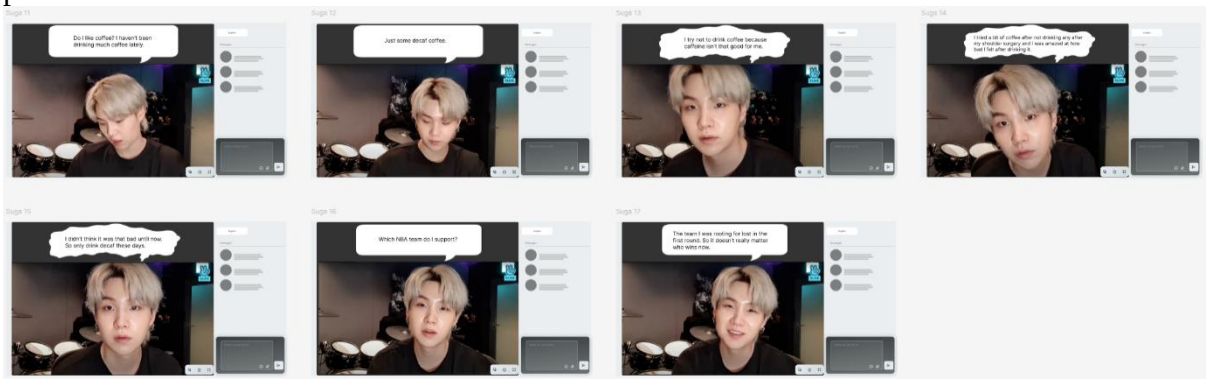


Figure 9. Figure Gridview8. Grid view, Suga's Story 2 (Bubbles)

3.2.3 Finalization of Study Session

We finalized that the study session's length was 45 minutes to consider technical scenarios that could come up considering it being conducted online through Teams meeting. We maintained that Abbas would be the facilitator while Truong the note taker. Truong would also be responsible for sharing screen to display the story, announcing the start and end of story as well as showing story grid view and bubbles design.

There was a change in design where the colour schemes were modified to shades of grey, black, and white. We anticipated that a relatively monochrome and mute theme would fully turn participants' attention onto the intended study design. The final design was shown in Figure X in Section 3.1 Experimental Designs. To increase randomization, we had an additional two study sets, Set 3 and Set 4, where the Control story starts first for both streamers. We also added in an extra panel used for introduction purposes as shown in Figure 10.

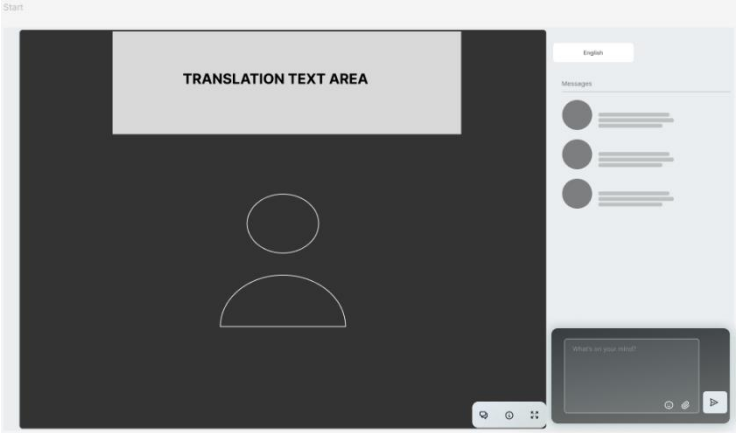


Figure 10. Intro Panel

Of no particular reason, the order of the study sets to use were Set 4, Set 1, Set 3 and Set 2. Out of all potential candidates, four participants were able to take part which coincidentally happened that each participant would have a unique study set. Our final preparation included designing a template panel to start as the very first frame. This decision aimed to introduce

participants to the general design so they could expect where to view the translation text area and the streamer.

3.2.4 Data Collection and Analysis

Data was collected by documenting our observations and each participant's answers during the study sessions. We also recorded each session so that we could rewatch the study and document any observations that we may have missed. We decided to rewatch the study sessions individually to limit bias put on one another. This allowed us to document our own observations, introducing different perspectives.

4 RESULTS

The results taken from our study are a combination of all four participants, which we will refer to as P1, P2, P3, and P4. It is important to note that participants P1, P2 and P3 attended or had recently graduated from university. They all expressed interest in the K-pop community and stated that they watch K-pop live streams or live-streams that have been uploaded to YouTube with translations. P4 is in her 50s and less involved in the K-pop community. However, she does watch streams that have been uploaded to YouTube. She expressed that she was a huge fan of Korean Dramas and relied on translations to keep up with her favourite shows. It is important to note that English is her second language, and she needed translation assistance with the study.

As stated, we asked the same three questions in between each story. Question one asked what emotions were expressed in each story, P1 seemed to pick up on the first emotion she found to be expressed in the stream. However, P2, P3, and P4 confidently expressed a multitude of emotions, picking up on the facial reaction of the streamer as well as which emotive translation bubbles were used in the story. For example, in the case of streamer Ryujin, story 3, P1 stated that the emotions she gathered from this were that of reminiscence. P2 expressed that the emotions included were thoughtfulness, excitement, joy, and energy. P3 stated that not only were excitement and happiness reflected in this story, but the emotive translation bubbles also made her feel those emotions. P4 expressed that she truly agreed and felt like the streamer when watching this story. This was a common trend throughout the study. With the exception of P1, participants could pick up on a multitude of emotions when viewing stories in which emotive translation bubbles were used.

However, when viewing the control case, participants relied more on the streamer’s facial expression and body language to catch the expressed emotion. The responses for question one were not as confident as when emotive translation bubbles were used. An important observation was that P3 and P4 did not comment as much on how the story made them feel as they did when the bubbles were used. P4 stated that the control story was dry and lacked emotions. P2 seemed a bit confused and unsure of which emotions were truly expressed. She stated that she had to rely on the context of the translation to pick up on their emotions.

Question two, for between stories, asked if the translation area matched the context of the stream. All participants unanimously agreed. P2, P3, and P4 pointed out that the translation area with the emotive translation bubbles made them feel more engaged in the stream than the ones without. P1 thought the control case was just a different type of emotive bubble. She

expressed that the emotion behind this was neutral.

After each streamer, we asked the participants to comment on how much the emotive translation bubbles impacted their understanding on a scale of 0 to 5, with 5 being extremely helpful. All participants responded with ranges of 4 to 5 for each streamer. An important observation was that participants stated that having different types of bubbles for different emotions was helpful as they could mostly skim the translations and still understand the context of the stories. P4 mentioned that she would find it even more helpful if emojis were used to express more emotions as she enjoyed reading comic books. All participants found the control bubbles to be bland and found they had to read more when viewing the control story.

Question two asked if participants were able to track the translations as well as the video scene. Although all participants stated they could do so easily, P1 and P4 mentioned that having the translation text box at the top of the screen was awkward instead of below the streamer. P1 mentioned that the "grey bubbles," the control case, were easier to read due to the decrease of contrast as opposed to the "white bubbles," the emotive translation bubbles. She also mentioned that if the text area is too large, it could detract from the streamer.

At the end of the study, we asked a series of questions to each participant. The first two questions were intended to investigate how the use of emotive translation bubbles affected the user's understanding and experience while watching the streams. All participants enjoyed using emotive translation bubbles and said it was helpful. P2 stated,

"*If it is live translation where there is grey background and no change in bubbles or what being emoted, it'd be consistently harder to perceive emotions being said. The streamer's facial expression doesn't explicitly match what they're saying. Having different speech bubbles would make it easier to tell consistently what emotion is being told. The speech bubbles Helps connect everything, so nothing gets missed.*"

P3 and P4 reached the same consensus. P4 stated that the emotive translation bubbles make watching the stream more engaging and she enjoyed that she could connect what she's reading to which emotions are expressed simultaneously.

"Definitely makes it more fun. You could see and read what's happening at the same time" – P4's response to question 1

"It will definitely make it better. Makes me more interested in the subject" – P4's response to question 2

P3 expressed that she would find this quite beneficial, making her enjoy the stream more.

"*Especially that last bit where there is a heart. Or when it changes when they're excited or sad. I feel that's very cute. It makes me more focused on the stream… I feel like if you can't fully catch what is in the speech bubble, whatever shape it was would make you understand it better." – P3.*

All participants agreed when asked if the emotive translation bubbles would impact the sense of connection to the idol. P2 felt more involved in what they were watching and connected to the idol, she stated that the squiggly line bubble and the heart bubble "pop off the screen more." Similarly, other participants expressed that the emotive bubbles help to connect to the idol because those shapes make them feel the same way the idol feels. P3 stated, "*You*

also get excited and feel like you can understand and connect with them better.".

Table 2 (Emotive Translation Bubbles & Consensus) displays what each participant thought the different bubbles meant. Although the participants had similar ideas of what each bubble meant, there was a slight difference among them.

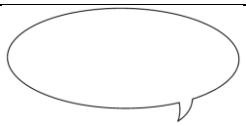
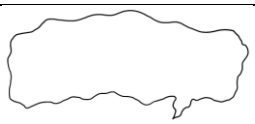
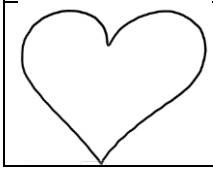
Bubble	P1	P2	P3	P4
	Looks like what a person would say verbally.	Flat sentence. If a question was asked and streamer was just repeating it.	Talking generally. Hi, Bye.	Official subject of conversation.
	Looks like it would be used for thoughts.	More of an excited, personal statement or a question.	A reply to a question they may have been asked. Or they are asking a question.	A normal day conversation.
	A combination of the first 2. She was not sure.	A loud statement. Something exciting or an announcement. If they're having a contest that day, for example.	Looks like an answer.	Not sure, maybe fear. Like they are scared or shocked.
	Looks similar to thought bubbles.	A more emotional statement. Like they felt really upset. Or could be excited with more emotion.	Confusion.	Like a map. Participants like this bubble and finds it stands out.
	Looks like something they would use if they were excited.	Used to express when you love something.	Used to express when they love something. Like their fans.	Used to express something they love.

Table 2. Emotive Translation Bubbles & Consensus

The study shed light on many interesting findings. One was how well three out of four participants responded to the emotive bubbles. They provided strong opinions on the bubbles that have pronounced styling, namely, 4 and 5. Participants reacted more neutrally towards bubbles 1 and 2 which have the familiar style that can be easily found in comic strips. Coincidentally, bubble 3 received mixed reactions when designed to express mixed or ambiguous emotions. Except for bubble 1 and bubble 5 where participants could clearly perceive their intention designs, the other bubbles received different interpretations.

The next finding was inspired by P2's feedback on creating a bubble resembling a cloud shape and P4's suggestion of including emojis to enhance the emotions further. The finding made us believe participants were more impressed by bubbles with distinct designs. The more outstanding the bubbles, the easier participants could comprehend and the more reactions they provided. While we continue to believe that a neutral statement should have a neutral and universal design, statements that denote emotions should have a very expressive speech bubbles design.

5 DISCUSSION

5.1 Reflections

Re-watching the interviews helped us reflect on a mistake that could have been avoided and revealed some minor adjustments that could have improved the study session experience. For example, on a few occasions, we forgot to display the grid view (specifically for P3 and P4). Even though no participants commented explicitly on the grid view being helpful during the actual user study sessions, it remained a step in the study that should happen for all participants regardless to remove inconsistencies. This experience is a good reminder to keep a technical checklist on the side. We should refer to the list constantly to ensure we follow the correct procedure.

One adjustment we thought of occurred when we switched between stories, Abbas would notify participants, and Truong would acknowledge and prompt the change by stating: “Story X starting now,” “Starting now” and so on. Even though we did not receive feedback that the story's start might have been too abrupt, we were not confident whether we provided participants enough time to mentally prepare. We would like to note that a transition panel or some on-screen cue to signify the starting of a new story and its ending might be worth considering. The transition panel or starting cue would help the participant shift their attention back to the UI after a series of questions. Regarding the end of the story’s need for a cue, when the last panel occurred, Sky would wait a few seconds to let the participants read the panel before notifying them. Nonetheless, there could be misjudgments or inconsistencies for the given time among participants; adding an ending cue here could improve an overall smooth and consistent experience.

One feedback we received from P4 was that she found that panel Ryujin 7 in Ryujin’s Story 1 (Control set) switched too fast for her to read. This was interesting since there were other panels with similar or even longer text length in Ryujin’s Story 2 and 3. In instance, panels Ryujin 17, Ryujin 23 and Ryujin 25 to name a few all had very similar text length to panel Ryujin 7, and they all had a six second switch time. However, P4 did not raise her concerns regarding the panels with speech bubbles. On one hand, the speed issue affirmed that we should have done deeper research into how many words a person can read on average and derived proper time. On the other hand, it could be a coincidence or an indication that P4 was able to follow emotive speech bubbles better.

5.2 Surprises & Limitations

As stated in our results section, the overall consensus was positive in this study. We were surprised to discover how much users relied on the emotive translation bubbles. Although, as noted, this could be due to us showing still images as opposed to videos. It would be interesting to see if the same consensus is reached if we were to use live videos, where users can use the streamer’s body language and tone to interpret emotion.

Another surprise was how positive the reactions were to speech bubble 5 (the heart). Without fail, when each user saw that shape on the stream, they reacted positively. This could be because that shape is universal for love, which is bound to have a positive reaction.

We understand that the emotive bubbles did not cater to all participants. To weigh in P1’s experience, P1 found the emotive speech bubbles distracting and the control, which P1 referred to as “grey speech bubbles,” to be more pleasant to focus on. It could be that users similar to

P1 would prefer a more subtle bubbles design, a different supplementary technique design, or simply a vanilla UI. Nonetheless, we took an interest in P1's comment: "It was obvious that without the translation area, there would be no content from the streamer." The comment sparked an idea in our minds: "What effects and to what extent can emotive speech bubbles have on user experience in accessibility mode?" We would like to think that users relying on purely texts and translation would benefit greatly from a supplementary technique like emotive translation bubbles, but that would require its own research to ascertain.

An important limitation to consider is that our UI design process did not need to follow standard UI design procedures. This was mainly due to the time constraints and our goals during this term. Our ultimate focus was understanding how emotive translation bubbles could potentially impact users' connection and understanding of a streamer. When deciding which UI to continue with, we had reservations about how well each would do. We anticipated more constructive feedback regarding the UI design and the emotive bubbles. However, we were surprised by how positive the reviews were. It is important to note that computer science researchers tend to look at these designs and studies with a more focused lens. The users themselves want designs that meet their goals and are "good enough." This gives us the confidence to move forward with future work.

5.3 Future Work

Regarding the next steps, this study has opened many doors for us to explore. One of them is following through on designing UI for emotive translation bubbles that follow proper design procedures. This could also entail developing a study that involves using live videos to explore the impact of these bubbles further. In terms of UI design, an eye tracker could be beneficial as it could help to determine the best possible designs to improve user-idol interactions.

Secondly, further exploration is needed to improve the accuracy of live multi-lingual translations. As noted, for the purpose of this study, we used accurate translations due to time constraints. Further work could build upon our previous NLLB study. Some recommendations include developing or integrating a model to transcribe live audio into the text and use the NLLB model to generate text-to-text translations. Another could be to incorporate the NLLB translations into a user study similar to this to determine if the results could be matched using a translation model.

One potential work involves determining what implications incorrect translations or emotive bubbles would produce. Would the user rely more on the emotive bubbles or on the translations themselves? This would be an interesting topic to explore as this study did not explore the edge cases.

The last suggestion can only be implemented if the study is done in person. We suggest using eye tracker to assess where on the screen the users pay the most attention to. From our study, we recognize that some users care more about streamer's facial expressions while some pay close attention to the translation and rely on the translation to establish understanding and connection. It is intriguing to investigate if there is a relationship between what users physically look for and how a UI affect their perception.

6 CONCLUSION

Automatic translation has become an essential tool in today's globalized world, where people from diverse cultures and languages interact with each other more than ever before.

However, machine translation could be better and often produces suboptimal results, which can lead to misunderstandings and misinterpretations. Hence, there is a need for improvement in machine translation, especially in user-centred design principles.

We focus on the K-pop fan inclusive community connects fans from various cultural and linguistic backgrounds worldwide. K-pop idols frequently interact with their fans through social media live streams, which can be challenging for fans needing help understanding Korean. In such scenarios, emotive translation bubbles representing the idol's emotions can effectively enhance the user's understanding and emotional connection to the content.

This informal study found that adding emotive translation bubbles significantly improved users' emotional connection and understanding of the content. This research highlights the potential benefits of incorporating user-centred design principles in improving machine-translation output. Small design changes, such as emotive translation bubbles, can make a significant impact on users' experiences, especially when it comes to language barriers.

As technology evolves, incorporating user-centred design principles in machine translation can lead to more inclusive and engaging content for all. The findings of this study can be applied to other contexts where language barriers pose a challenge in understanding and connecting with content. By leveraging design thinking principles, designers and developers can create more effective and engaging machine translation solutions that cater to users' needs and preferences. Ultimately, this will lead to more effective communication and better cultural understanding between people from different linguistic and cultural backgrounds.

REFERENCES

[1] H. H. Clark and S. E. Brennan (1991). Grounding in communication. *Perspectives on socially shared cognition*. L.B. Resnick, J.M. Levine, and S.D. Teasley, eds. American Psychological Association. 127–149.

[2] Laila Homaeian. Supporting Communication in Machine-Translation (MT) Systems: A Literature Review.

[3] Robertson, S., & Díaz, M. (2022). Understanding and Being Understood: User Strategies for Identifying and Recovering From Mis translations in Machine Translation-Mediated Chat. ACM FAccT. pp. 2223-2238

[4] Lim, H., Cosley, D., & Fussell, S. R. 2022. Understanding Cross-lingual Pragmatic Misunderstandings in Email Communication. 6(CSCW1), 1-32

[5] Lim, H., Cosley, D., & Fussell, S. R. 2018. Beyond translation: Design and evaluation of an emotional and contextual knowledge interface for foreign language social media posts. CHI. pp. 1-12.

[6] Gao, G., Xu, B., Hau, D. C., Yao, Z., Cosley, D., & Fussell, S. R. (2015). Two is better than one: improving multilingual collaboration by giving two machine translation outputs. CSCW. pp. 852-863

[7] Shigenobu, T. 2007, July. Evaluation and usability of back translation for intercultural communication. UI-HCII. Lecture Notes in Computer Science, vol 4560. Springer.

[8] Miyabe, M., & Yoshino, T. (2009). Accuracy evaluation of sentences translated to intermediate language in back translatio n. In Proceedings of the 3rd IUCS. pp. 30-35.

[9] Gao, G., Wang, H. C., Cosley, D., & Fussell, S. R. (2013). Same translation but different experience: the effects of highlighting on machine-translated conversations. CHI(pp. 449-458).

[10] Pan, M. H., Yamashita, N., & Wang, H. C. (2017). Task rebalancing: Improving multilingual communication with native speak ers-generated highlights on automated transcripts. CSCW. pp. 310-321.

[11] Pan, M. H., & Wang, H. C. (2014). Enhancing machine translation with crowdsourced keyword highlighting. CABS. pp. 99-102.

[12] Green, S., Heer, J., & Manning, C. D. (2013). The efficacy of human post-editing for language translation. CHI. pp. 439-448.

[13] Hara, K., & Iqbal, S. T. (2015). Effect of machine translation in interlingual conversation: Lessons from a formative study. CHI. pp. 3473-3482.

[14] Szarkowska, A., Silva, B., & Orrego-Carmona, D. (2021). Effects of subtitle speed on proportional reading time: Re-analysing subtitle reading data with mixed effects models. *Translation, Cognition & Behavior*, *4*(2), 305-330.

[15] T. Aoki, R. Chujo, K. Matsui, S. Choi and A. Hautasaari. EmoBalloon - Conveying Emotional Arousal in Text Chats with Speech Balloons. Proc.Conference on Human Factors in Computing Systems (CHI). 16, Article 527, 1–16, 2022. https://doi.org/10.1145/3491102.3501920

[16] Animeoutline. (2020, July 4). How to draw manga speech bubbles tutorial. AnimeOutline. Retrieved April 22, 2023, from https://www.animeoutline.com/how-to-draw-manga-speech-bubbles/

[17] Hernández, I. P. (2022, July 17). Meaning of speech bubbles in comics. Ilka Perea Studio.

Retrieved April 22, 2023, from https://ilkaperea.com/2019/08/15/meaning-of-speech-bubbles-in-comics/

[18] Arcoya, E. (2020, November 27). Comic speech bubbles: What are they and what types of balloons are there? Creativos Online. Retrieved April 22, 2023, from https://www.creativosonline.org/en/bocadillos-de-comic.html?msclkid=fef25de5b9b811ec9fab2a9ab8191644

[19] YouTube. (2021). [Eng Sub] Bts Suga Live Vlive (2021.07.19) Bts Vlive 2021. YouTube. Retrieved April 22, 2023, from https://www.youtube.com/watch?v=QtqG16sxqMY&t=906s&ab_channel=BTSLIVE.

[20] YouTube. (2021). (Eng/Indo/Jpn Sub) Itzy Ryujin Vlive 211104 | [#ryujin] Wawm 😉. YouTube. Retrieved April 22, 2023, from https://www.youtube.com/watch?v=4nd5VXSpRKI&ab_channel=KLive.

APPENDIX A:

Study Design

Main Research Question:

Does incorporating the speech bubble user interface for live K-Pop streams improve the understanding and experience between the idol and user?

Associating Questions:

- Does the use of a speech bubble user interface during live K-Pop streams result in higher levels of engagement and interaction between the idol and the user, as measured by comments, likes, or shares?
- Are users who are given a speech bubble user interface during live K-Pop streams more likely to feel connected to the idol and the performance compared to those who aren't given the interface?
- Does incorporating a speech bubble user interface during live K-Pop streams lead to higher levels of satisfaction and enjoyment among users?
- Are there any significant differences in the understanding and experience of live K-Pop streams among users of different ages, genders, or nationalities when a speech bubble user interface is incorporated?

Variables:

- Dependent:
    - The user's experience, including levels of engagement and understanding.
- Independent:
    - The user interface (speech bubbles) for translation

Proto-Persona:

Beck Quinn

- Age: 20
- Gender: Female
- Nationality: Canadian
- Occupation: Student
- Education: Majoring in Psychology
- Proficiency: Fluent in English; No knowledge of Korean

Goals and Needs:

- Beck is a huge fan of K-pop and watches live streams of their favorite groups regularly.
- They want to understand their idols as they this helps them communicate and build a closer relationship with them.
- They don't know any Korean and rely heavily on translations during live streams.
- They value translations that are accurate, easy to understand, and provide context for the streams.

- As a psychology major, they are interested in understanding the social and cultural aspects of K-pop fandom.

Pain Points:

- Difficulty understanding the translations during live streams due to lack of knowledge of the Korean language.
- Inaccurate translations that don't provide context for the streams.
- Limited engagement with idol during live streams.
- Limited knowledge of Korean culture and language, which makes it challenging to fully appreciate the idols and emotions behind their words.

Participant Recruitment Procedure

2. Reach out to people who are interested in participating
3. Ask the participants the questions below; if the responses do not fit based on the proto-persona, do not proceed.
    1. Q: Do you watch idols live stream?
    2. Q: Do you interact with idols on the live stream?
    3. Q: Do you use translation applications/models to further immerse yourself in the K-Pop culture?
4. Ensure they fit the proto-persona based on the description below
5. Begin the study

How participants align:

The ideal participant for this recruitment procedure is someone who follows our proto-persona. Our first question ensures that the participant has to an extent some knowledge of the K-Pop industry. The second question ensures that the participant may have interest in participating in live streams with their idols, despite language barriers. The last question indicates whether the participant will be an ideal candidate to use translation models to immerse themselves within the community.

Preparatory Setup Steps:

1. Create a MS Teams meeting link, that participants and team members can use to join.
2. Setup a recording consent form if the participant agrees to have the session recorded
    a. Have this signed prior to the session.
3. Setup MS Teams document for taking notes
    a. For each participant, create a template that consists of a section for:
        i. Overall observations during the study.
        ii. Template listing the tasks the user will be asked to complete & a section for their answer of how the process was, as well as their rating of the tasks.
        iii. A list of follow-up questions and their answers.
        iv. Three questions were developed to test the hypothesis.
4. Record the participant number.
5. Have the Figma sharing link ready for participants.

Study Script:

1. Introduction & explain the purpose:
    - "Hello and thank you for agreeing to participate in our user observational study. During this study, we will be showing you different K-pop live streams with the speech bubble user interface prototype. We will ask you to watch the live streams and provide feedback on your understanding of the streams, engagement, enjoyment, and overall satisfaction with the live stream. Your feedback will help us determine if adding the speech bubble user interface to translations improves the user experience. Please be assured that your feedback will be kept confidential and will only be used for research purposes. If you have any questions at any time during the study, please feel free to ask. Thank you again for your participation."

2. Reiterate the participation consent emailed prior to:
    - “Before we continue, we would like to make sure you are happy to continue and participate in the study. As a reminder, when agreeing to participate in this study, you have answered the following questions.”
        - Q: Do you watch idols live stream?
        - Q: Do you interact with idols on the live stream?
        - Q: Do you use translation applications/models to further immerse yourself in the K-Pop culture?
    - “We now ask that you have your camera turned on so we can observe your reaction throughout the study.”

3. Explain that there will be a note-taker as well as other roles:
    - “In this session, **Zayn** will be a facilitator who will communicate the steps and help facilitate the session. **Zayn** will be the prototype session leader, who will provide you with tasks and ask you to complete them with the prototype you will be provided. **Sky** will be the note taker, and they will write down key points they notice as the study session continues. **Sky** will be the observer and will assist in keeping track of observations of how the process is going.”

4. Ask: “Do you have any questions regarding the testing session?”

5. Ask: “Are you comfortable sharing your screen?”
    - If they are, the helper must share the Figma link with them through the Zoom chat.
        i. Ask the user: “Can you screen share the Figma Prototype? Please do not click anything until we tell you to.”
    - If not, open the Figma link on the facilitator’s computer and share your screen.
        i. Give participant control of the facilitator’s browser
6. “Are you ready to begin?”

7. “We will now begin the experiment; the prototype session lead, **Zayn** will guide you through this process and ask you to complete tasks. There will be a number of tasks that you are expected to conduct, this session’s helper, **Zayn,** will also assist you during the process. If you have any questions or are confused, please feel free to stop the task and ask the helper.”

8. Show participant translation area, idol, and message section [Explain the UI]

9. The note taker must record the date and start time.

10. “Please remember to think out loud during this session. We will ask you to complete a task; then, we will follow up by asking about your experience and rating its difficulty. Feel free to explain your thought process as you go or anything you notice. This could be something you find interesting, confusing, easy, or hard to use.”

11. Tasks [For each stream, ask these four questions]:
    - “We will now show you some streams and ask you questions pertaining to those streams.”
    - Q1: What was your understanding of the stream?
    - Q2: How much did the speech bubbles improve your understanding on a scale of 0 to 5?
        i. 0 is not helpful, and five is extremely helpful.
    - Q3: Was it difficult to track the translation and listen to the stream simultaneously?
    - Q4: How accurately did the speech bubbles reflect the emotion of the streamer?

12. Ask research questions:
    - “Now that we have finished the streams, we will ask you some questions pertaining to the study as a whole.”
    - Q1: Can you describe how the use of a speech bubble user interface during live K-Pop streams affects your enjoyment of the stream?
    - Q2: Can you describe how the use of a speech bubble user interface during live K-Pop streams affects your understanding of the stream?
    - Q3: Does the use of speech bubbles have an impact on how connected you feel to the idol?

Post-Participant Instructions:

- Add to the MS Teams doc any last observations or notes from the study session the notetaker/observer made
- Add details to the notes for any missing or unexplained thoughts
- Discuss feedback received for the study session and incorporate any constructive feedback that could make the process smoother for the next session [Ok for pilot study]
- If the participant made recommendations, record that.
- Reset Figma prototype

Study Template

Date: **Start Time:**

Participant Recruitment Procedure Questions / Answers

1. Q: Do you watch idols live stream?
2. Q: Do you interact with idols on the live stream?
3. Q: Do you use translation applications/models to further immerse yourself in the K-Pop culture?

In between Stories:

Q1: What were the emotions expressed in this story?
Q2: Does the translation widget accurately match the content?
Q3: Do you have any other comments?

End of Streams:

Q1: How much did the speech bubbles impact your understanding on a scale of 0 to 5? (0 being not helpful, five being extremely helpful.)
Q2: Were you able to attend to the translation as well as the video scene?
Q3: Are there any other comments you have on this stream?

Any Questions the Participant Asked + Answers:

Post-Experiment Questions:

- Q1: Can you describe how the use of a speech bubble user interface during live K-Pop streams may affect your experience during the live stream?
- Q2: Can you describe how the use of a speech bubble user interface during live K-Pop streams may affect your understanding of the stream?
- Q3: Could you describe whether the use of speech bubbles would impact your sense of connection to the idol?
- Q4: What current strategies do you use to understand live-streaming content?
- Q5: Display each speech bubble and ask them to comment on what they think they represent.
- Q6: Do you have any feedback about the design of speech bubbles?
- Q7: Do you have any feedback on the study session?
- Q8: Do you have any other comments?