FaceType

Crafting Written Impressions of Spoken Expression

KEVIN MAHER, Tsinghua University, China

FAN XIANG, Tsinghua University, China

LIANG ZHI, Tsinghua University, China

FaceType is an interactive installation that creates an experience of spoken communication through generated text. Inspired by Chinese calligraphy, the project transforms our spoken expression into handwriting. FaceType explores what parts of our spoken expression can be evoked in writing, and what the most natural form of interaction between the two might be. The work is aimed to allow lay audiences to experience emotion, emphasis, and critical information in speech. Further audience reflection about patterns in their expression and the role of unconscious and conscious expression provide new directions for further works.

CCS Concepts: • **Applied computing** → Arts and humanities; Fine arts; • **Human-centered computing** → Human computer interaction (HCI); Interaction paradigms; Natural language interfaces.

Additional Key Words and Phrases: Text visualization, Artistic visualization, Visual communication

1 INTRODUCTION

Our work began with curiosity about what makes our speech personal and unique. In our everyday speech, how something is said has a significant impact on how information is communicated. Facial expression, speech tone, and word speed all are critical to understanding a speaker. In our search for how to communicate this information, we found that possibilities for expressing and interacting with this information were not fully explored.

In the case of digital type, this is especially true. Digitally typed words lose critical information about the original emphasis and emotional impact of a speaker. While the popular misuse of Mehrabian's research that only 7% of communication comes from the meaning of the words we say is disputed [Lapakko, 2007; Yaffe, 2011], beyond the words we say, there is important communication that is expressed but left out of digital recordings. What could be revealed about us once our emphasis, emotion, and speech patterns are recorded?

FaceType began as an exploration to evoke this information in an intuitive way to audiences. This study shows the development of a visual system – FaceType, that takes emotional expression and vocal emphasis as a data source and generates text visualizations based on Chinese calligraphy. Interaction with FaceType is meant to evoke interest and reflection about our natural speech.

We found inspiration for FaceType in the expressive nature of calligraphy from China and Japan, where we saw the forms of the written word used expressively, often with emotion and emphasis

38

Authors' addresses: Kevin Maher, Tsinghua University, China, kevintmaher@gmail.com; Fan Xiang, Tsinghua University, China, xiangfan@mail.tsinghua.edu.cn; Liang Zhi, Tsinghua University, China, zhil19@mails.tsinghua.edu.cn.

Fig. 1. Detail of "Draft of a Requiem to My Nephew" by Yan Zhengqing. Photograph by Taipei: National Palace Museum [Public Domain] via Wikimedia Commons. (https://bit.ly/33umbJm)

given a visual form. In the work "Draft of a Requiem to My Nephew" crafted by Yan Zhengqing in 758 AD, we can see the author begin with a formal and restrained script on the top right (Figure1). The brush strokes quickly morph into more unrestrained emotional turmoil as he writes about his deceased nephew.

The dynamic nature of speech can be well reflected in rapid changes in Chinese calligraphy. Ebrey summarizes "most European calligraphy is highly stylized, regular, and decorated with flourishes, which in themselves are lacking in personal expression [it tends] to curb spontaneity, producing fairly static forms." [Ebrey, 2001] In contrast, in Chinese calligraphy, the importance of the process of creation, personal expression, and visual emphasis of characters made the medium an appealing visual metaphor for our work.

2 RELATED WORK

From our survey of technical works, we recognize that there is an undeveloped potential for representing information about our communication through text. Prosodic font was developed to generate text that reflects the pitch and rhythm of speech [Rosenberger and MacNeil, 1999]. A survey of text visualization methods by Brath et al. notes that prosodic font is the only case of changing glyphs of fonts to express quantitative data [Brath and Banissi, 2016]. Recently new methods of generating fonts that could be then used to encode non-verbal are emerging. Wang created Attribute2Font which allows for creation of new fonts based on attributes, including emotions [Wang et al., 2020]. While Chen et al. used methods referencing graphology [Chen et al., 2018], the study of how handwriting can reveal one's personality, we found the scholarly writing on graphology controversial. Our work more narrowly focuses on where handwriting intentionally or unintentionally expresses emotion, especially calligraphy.

In artistic works Golan Levin developed Ursonography which used real-time speech recognition to show a performed sound poem. A goal of our work, similar to Levin's is to create "new forms of interactive media in order to explore novel conduits for non-verbal expression and communication" [Levin, 2013]. In contrast to Ursonography, which focused on sound, we were interested in showing emotional expression, which in fact can be the root cause of sound properties of a voice.

Fig. 2. The process of mapping emotion to calligraphy stroke. (©Kevin Maher, 2022)

In our previous work, E-ffective: A Visual Analytic System for Exploring the Emotion and Effectiveness of Inspirational Speeches, we developed typographic methods to visually communicate word speed, pause, and emotional data [Maher et al., 2022]. This work provided the technical foundation of FaceType, as it both provided familiarity with the different possibilities of data available, including emotion contained in the voice, words said, and facial expression, as well as various visualization technologies. However, the purpose of the work and visualization methods were entirely different, as it was an analytical tool for public speaking delivery techniques, with the visualization method of Tufte's "data ink ratio" maximized [Tufte, 2001]. Instead, FaceType's purpose is open-ended exploration, with our "data ink" crafted as a visual metaphor, imperfect and subjectively devised. This is in keeping with the work of many calligraphers who often are indifferent to legibility and style consistency.

3 INSTALLATION PERFORMANCE

3.1 Emotional Expression

We were most curious about the role of speakers' facial expression during speech. The author of this work had personal curiosity about the nature of our expression, having lived in both the United States and China, he experienced very different emotional expressions. If recorded, could different emotional patterns of speakers be observed and expressed intuitively?

We wanted to be able to see emotional nuances of speakers, rather than emotional categories such as "happy" so we used the emotional model provided by Mollahosseini et al. that estimates the relative emotional intensity and degree of positivity or negativity of a face [Mollahosseini et al., 2019]. To display this emotional information in a way similar to the craft of calligraphers, collaboration was made with a graduate student majoring in Chinese calligraphy. With historical samples as a reference, as well as her personal opinion, samples of "emotional brush strokes" were made, as seen in Figure2. Attributes of calligraphy such as brush dryness, form, and animated action were refined, patterned off the process of writing Chinese calligraphy. The suitability of this mapping method, shown in Figure3, was supported by other practicing calligraphers.

To apply the emotionally spontaneous and dynamic nature of calligraphy in "brush strokes" we further developed the open source font Leon Sans [Kim, 2019] displaying the emotional brush strokes along the paths of each English letter. For example, if the facial emotion of an audience member was very positive, strokes that were more curved and rounded were applied to the font. If the emotional intensity of the audience was high, strokes of the letterforms appear to have more
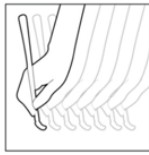
Fig. 3. Relationship between emotion and brush attributes. (©Kevin Maher, 2022)

"ink". In this mapping method, more intense emotions and emotions that were on either extreme of the positivity and negativity spectrum were highlighted. We wondered if our emotional expressions were infused in the letterforms in English in a way similar to Chinese calligraphy, how would our understanding of our speech be changed?

3.2 Expressing Emphasis

The spontaneous and dynamic nature of calligraphy can reveal emphasis of words. Similarly, we crafted ways of encoding critical speech data. Natural language processing provided English voice speech to text as well as pause, speed, and syllable data. A first step was to show the same kind of emphasis that occurs in Chinese calligraphy. Spoken word speed, calculated by the time is takes to say each syllable in a word, was used as an indication of word emphasis, and indicated by the width of the generated fonts and spacing of the characters within a word. Pauses were naturally shown by the distance between words. As can be seen in Figure 4, the same word said with different emphasis and emotional intensity, results in different fonts and typography. In the original installation we chose character size to emphasize emotional intensity as well as word speed, a decision arbitrarily made given the importance we imagined of these two metrics.

3.3 Installation Design

The installation for FaceType was made for audiences to interact naturally with the work. The original installation was in China where calligraphy is more of a part of everyday culture, and is a more readily understood metaphor for expression. A prominent microphone and camera in front of the interface suggest the mode of interaction, as seen in Figure 5. As nearly half of the audience interacted with the work, the initial interaction can be seen as readily understandable.

Similar to Chinese calligraphy, writing on FaceType begins from top to bottom for each letter of the words said by the audience, and then continues for each new phrase from right to left. In
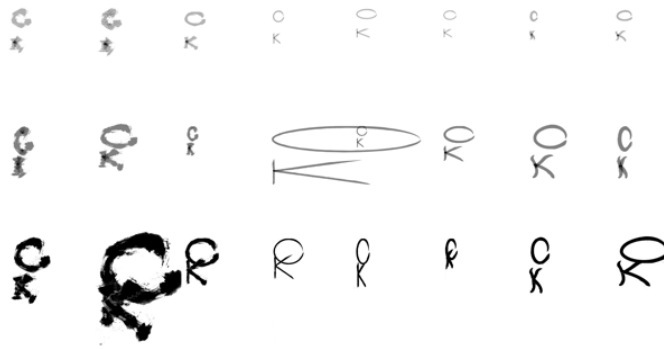
Fig. 4. Different expressions of the word "ok" using FaceType. (©Kevin Maher, 2022)

Fig. 5. Concept image of FaceType. (©Kevin Maher, 2022)

the original showing at the Tsinghua University Postgraduate Exhibition, the interface is divided into an LCD screen that shows the most recent audience recordings on the right, and less recent recordings projected onto a painting canvas on the left (Figure 6).

4   INTERACTION AND ITERATION

When audiences approached the installation for the first time, many appeared surprised at the responses of the visual forms. Some would attempt to change the look of their recordings to be more beautiful or regular, revising their expression as they became aware of their role in creating different visual forms. However, these audience iterations revealed problems that resulted in the corresponding iteration of our work, as well as our own understanding of spoken expression.

Since the first installation was shown in China, where many in the audience were not as familiar with the input language of English, the authors originally wondered if there would be difficulties

Fig. 6. FaceType as seen in the 2021 Tsinghua University Postgraduate Exhibition. (©Kevin Maher, 2022)

Fig. 7. Emotionally intense users with their generated speech. (©Kevin Maher, 2022)

with interaction, as verbal expression would be limited by non-fluent English. However the limited ability of some of the audience to verbally express themselves did not seem to limit audiences to express their emotion and emphasis. As seen in Figure 7, an audience with minimal ESL elementary vocabulary expressed themselves with powerful emotional expression and emphasis. Some users that had good command of the English language, attempted more challenging verbal expression such as acting out parts of Hamlet, or reading poetry.

There were difficulties by many in the audience to understand how to use their face as a mode of interaction. An audience member with a furrowed brow remarked "why does my text look so

Fig. 8. Use of FaceType by several users over an hour of time. (©Kevin Maher, 2022)

ugly?" apparently unaware of the negative expression on his face. Stark differences in the visual recordings by different audience members, as seen in Figure 8, often resulted in confusion about the cause.

We observed many first-time users assuming that sound was the primary form of interaction, as they would speak loudly into the microphone, then seem confused by the unaffected letterforms. We reflected that perhaps voice is a more natural form of interaction in speech, something the audience feels they can control, after all, in everyday conversation we can hear our own voice expression, but cannot see our own face. A study by Qu et al. found that participants have relatively low self-awareness of their facial expressions, which is strongly increased under the condition that they can see their own faces by video [Qu et al., 2017]. Perhaps the relative lack of awareness of one's own facial expression contributed to the difficulty of the interaction. Additionally, after the installation showing, in our review of the audience facial expression data we found that the system glossed over important emotional changes of the audience, as it sampled facial emotion of every sentence. Given than the audience may have expressed more than one emotion in one sentence, to have strong emotion averaged out by the system may have made the interaction feel inaccurate
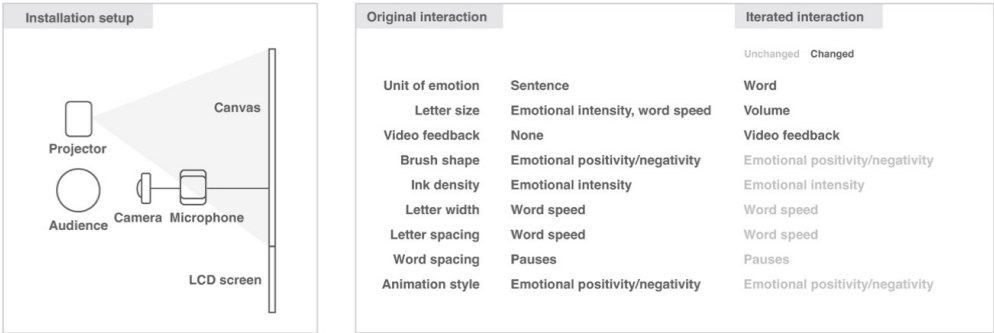
Fig. 9. A summary of the original interaction and changes made to FaceType. (©Kevin Maher, 2022)

Fig. 10. FaceType iterated. (©Kevin Maher, 2022)

and confusing. With these difficulties of interaction in mind the original installation was iterated, as summarized in Figure 9.

We changed the data structure of FaceType so that multiple emotions were sampled per word, reflecting audience micro-expressions more dynamically. Given that many users repeatedly attempted to use voice volume as a method of interaction, we added it as a method of interaction, and gave it a dominant role as the size of characters of the generated font. In later showings of our work, we observed this as a more intuitive way of interaction. To bring unconscious facial expression to the awareness of users, we added video feedback (Figure10). However, in later showings this presented a further challenge to the interaction as some people were camera shy, appearing concerned about how they would appear on camera. Thus using our face expression as means of interaction both presented challenges when used without video feedback due to lack of user awareness, as well as challenges when video was used, as it inhibited audience interaction for some users. By bringing faces more consciously into the piece, distractions are created that detract from the work.

5 CONCLUSION

FaceType was created with the visual metaphor of calligraphy in mind to express the dynamic nature of our speech. Similar to calligraphy, it was subjectively crafted and imperfect in nature. As

an open-ended exploration, it revealed aspects about the conscious control and understanding of our emphasis and emotions.

While the tradition of English calligraphy lacks the emphasis on dynamic emotional expression that Japanese or Chinese has, we found that a fusion of English letterforms and dynamically changing letter strokes and typography could result in more naturally understood traces of our expression. Facial expression, similar to calligraphy, can express subtle, delicate, and dynamic changes.

For future iterations by others seeking to craft written impressions of spoken expression, there are hints of new directions for work. The face is one of the most natural ways to interpret our emotion, data visualization expert Colin Ware claimed "although sound can be transmitted cheaply, sincerity, or the lack of it, is conveyed by subtle facial expressions, which is why face-to-face meetings are often worth the cost of travel" [Ware, 2022]. How to more naturally display and interact with critical information in our facial expression remains a challenge. Our spoken expression contains many different factors: gestures, voice pitch, speed, pause, volume, and more. In our everyday speech these factors show flavors of our personalities, if only technology could visualize them in a way that shows our subtle and unique signatures.

REFERENCES

Richard Brath and Ebad Banissi. 2016. Using typography to expand the design space of data visualization. She Ji: The Journal of Design, Economics, and Innovation. (Spring 2016), Tongji University Press 2, 1 59-87. https://doi.org/10.1016/j.sheji.2016.05.003

Ziqian Chen, Gentiane Venture, and Marie-Luce Bourguet. 2018. A rendering model for emotional in-air handwriting. Proceedings of British HCI (July 2018), BCS Learning and Development 1-5. http://dx.doi.org/10.14236/ewic/HCI2018.111

Patricia Ebrey. 2001. A visual sourcebook of Chinese Civilization:calligraphy. Course Materials, Department of History, University of Washington. https://depts.washington.edu/chinaciv/

Jongmin Kim. 2019. Leon Sans. GitHub repository. https://github.com/cmiscm/leonsans

David Lapakko. 2007. Communication is 93% nonverbal: An urban legend proliferates. Communication and Theater Association, (Dec. 2007), CTAMJ 34, 1 7-14. https://cornerstone.lib.mnsu.edu/ctamj/vol34/iss1/

Golan Levin. 2013. Seminar Series, Department of Computer Science, Carnegie Mellon University. https://www.cs.cmu.edu/~CompThink/seminars/golan/index.html

Kevin Maher, Zeyuan Huang, Jiancheng Song, Xiaoming Deng. Yu-Kun Lai, Cuixia Ma, Hao Wang, Yong-Jin Liu, and Hongan Wang. 2022. E-ffective: A Visual Analytic System for Exploring the Emotion and Effectiveness of Inspirational Speeches. IEEE Transactions on Visualization and Computer Graphics. (January 2022), IEEE 28, 1 508 – 517. https://doi.org/10.1109/TVCG.2021.3114789

Ali Mollahosseini, Behzad Hasani, and Mohammad H. Mahoor. 2019. AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. IEEE Transactions on Affective Computing. (January 2019), IEEE 10, 1 18-31. https://doi.org/10.1109/TAFFC.2017.2740923

Fangbing Qu, Wen-Jing Yan, Yu-Hsin Chen, Kaiyun Li, Hui Zhang, and Xiaolan Fu. 2017. "You Should Have Seen the Look on Your Face. . .": Self-awareness of Facial Expressions. Frontiers in Psychology. (May 2017), Frontiers 10, 1-8. https://doi.org/10.3389/fpsyg.2017.00832

Tara Rosenberger and Ronald L. MacNeil. 1999. Prosodic font: translating speech into graphics. In CHI '99 Extended Abstracts on Human Factors in Computing Systems (CHI EA '99). ACM 252–253. https://doi.org/10.1145/632716.632872

Edward Tufte. 2001. The Visual Display of Quantitative Information (2nd. ed.). Graphics Press, Cheshire, CT.

Yizhi Wang, Yue Gao, and Zhouhui Lian. 2020. Attribute2Font: creating fonts you want from attributes. ACM Trans. Graph. (August 2020), ACM 39, 4 1-15. https://doi.org/10.1145/3386569.3392456

Colin Ware. 2022. Visual Thinking for Information Design (2nd. ed.). Elsevier Science, Cambridge, MA.

Philip Yaffe. 2011. The 7% rule: fact, fiction, or misunderstanding. Ubiquity, (October 2011), ACM 1. 1-5. https://doi.org/10.1145/2043155.2043156