Designing voice-aware text in voice media with background color and typography

Qinyue Chen, Yuchun Yan and Hyeon-Jeong Suk*

Color Laboratory, Department of Industrial Design, KAIST, South Korea
**Email: color@kaist.ac.kr*

Speech-to-Text (STT) plays a significant role in voice media. While preserving semantic information, STT also results in a large loss of nonverbal information in the voice. The goal of voice-aware text design is to bridge the gap between expressive voice and its converted text. An online survey was carried out to compare the effect of different text design elements - font, background color, and typography - on emotion expressivity, content delivery, and appropriateness of converted text. Background color and typography enhanced all three scales; however, the font did not, and even had negative effects on content delivery and appropriateness. We also found that the combination of background color and typography was regarded as the most appropriate text design in both voice messaging and social media.

Reprinted version published online: 26 January 2022
Original source: Proceedings of the 14th Congress of the International Colour Association (AIC2021)

Introduction

Voice media, such as voice messaging and voice posts on SNS, have become increasingly popular. Many recent studies have focused on Speech-to-Text (STT), which is only concerned with the correctness of converted text. However, voice conveys more than contents when compared to plain text. STT inevitably results in a significant loss of nonverbal information in the voice that may have supported the user's emotional expression and content delivery.

To address this question, researchers began to use text design to represent voice characteristics and emotions. For example, studies have demonstrated the effectiveness of voice-driven typography in representing voice characteristics such as loudness, speed, and prosody [1]. Moreover, several researchers attempted to compensate for the emotion loss in mobile messaging by using affective fonts [2], color [3] or emoji [4]. However, it remains unknown which text design is most appropriate for visualizing voice media.

In this study, we were inspired to visualize voice media by changing the font, background color, and typography. An online survey was conducted to compare the effect of these three text design elements on emotion expressivity, content delivery, and appropriateness of converted text. We also investigated the most appropriate text design of converted text in voice messaging and social media.

Method

Voice stimuli

We selected voice stimuli from the CREMA-D [5] corpus that corresponded to utterances of the same content given with various emotions. The same actress read aloud all of the voice stimuli in six emotions: happy, surprised, sad, fearful, angry, and disgusting. The voice stimuli consisted of three short sentences and lasted around 8 seconds. By purpose, these sentences are emotionally neutral in their literal meaning, avoiding meaning hints from the textual content.

Voice-aware text design

The goal of this study was to investigate the effect of voice-aware text design in voice media. More specifically, three text design elements - font, background color, and typography, were investigated. We decided to map the text design to the voice in the following way:

First, as shown in Table 1, different fonts and background colors were selected to represent seven voice emotions. When the font or color change, the participants are supposed to be aware of the emotion change immediately and understand the target emotion intuitively.

Emotion	Font	Background Color	
Netural	Maybe tomorrow it will be cold. I would like a new alarm clock. I wonder what this is about.	Grey	#BFBFBF
Happy	**Maybe tomorrow it will be cold. I would like a new alarm clock. I wonder what this is about.**	Light blue	#01b0f0
Surprised	**Maybe tomorrow it will be cold I would like a new alarm clock I wonder what this is about**	Orange	#F5C243
Sad	*Maybe tomorrow it will be cold. I would like a new alarm clock. I wonder what this is about.*	Dark blue	#134d81
Fearful	Maybe tomorrow it will be cold. I would like a new alarm clock. I wonder what this is about.	Purple	#7030A0
Angry	Maybe tomorrow it will be cold. I would like a new alarm clock. I wonder what this is about.	Red	#C00000
Disgusting	Maybe tomorrow it will be cold. I would like a new alarm clock. I wonder what this is about.	Dark green	#0F460D

Table 1: Representative fonts and background colors for each emotion.

Font

Choi *et al.* [2] provides a 100-Font dataset where each font was labeled with six emotion categories (happy, sad, surprised, fearful, angry, and disgusting). We selected fonts with high emotional consensus in this dataset. For legibility reasons, we replaced several fancy or gothic fonts with other typefaces that have similar font characteristics. The font Arial is selected as a neutral font.

Background color

Multiple studies have confirmed that red is connected with anger [6-7]. Furthermore, people link light blue with happiness and pleasure [8], while dark blue is associated with sadness [6]. Orange is frequently connected with surprise, green with disgust, and purple with fear [7].

Typography

We mainly capitalized or slanted the word according to the acoustic features of voice stimuli, like word-level average loudness and speed. When we want to get a listener's attention, we always employ a louder voice in our speaking. Here, we use capitalized for the loudest words in a sentence to capture the reader's attention since it makes scanning the text and recognizing relevant keywords easier. Additionally, letter slant has been associated with changes in speed [1]. Thus, we map the voice speed to letter slant and utilize slanted for fast words in a sentence.

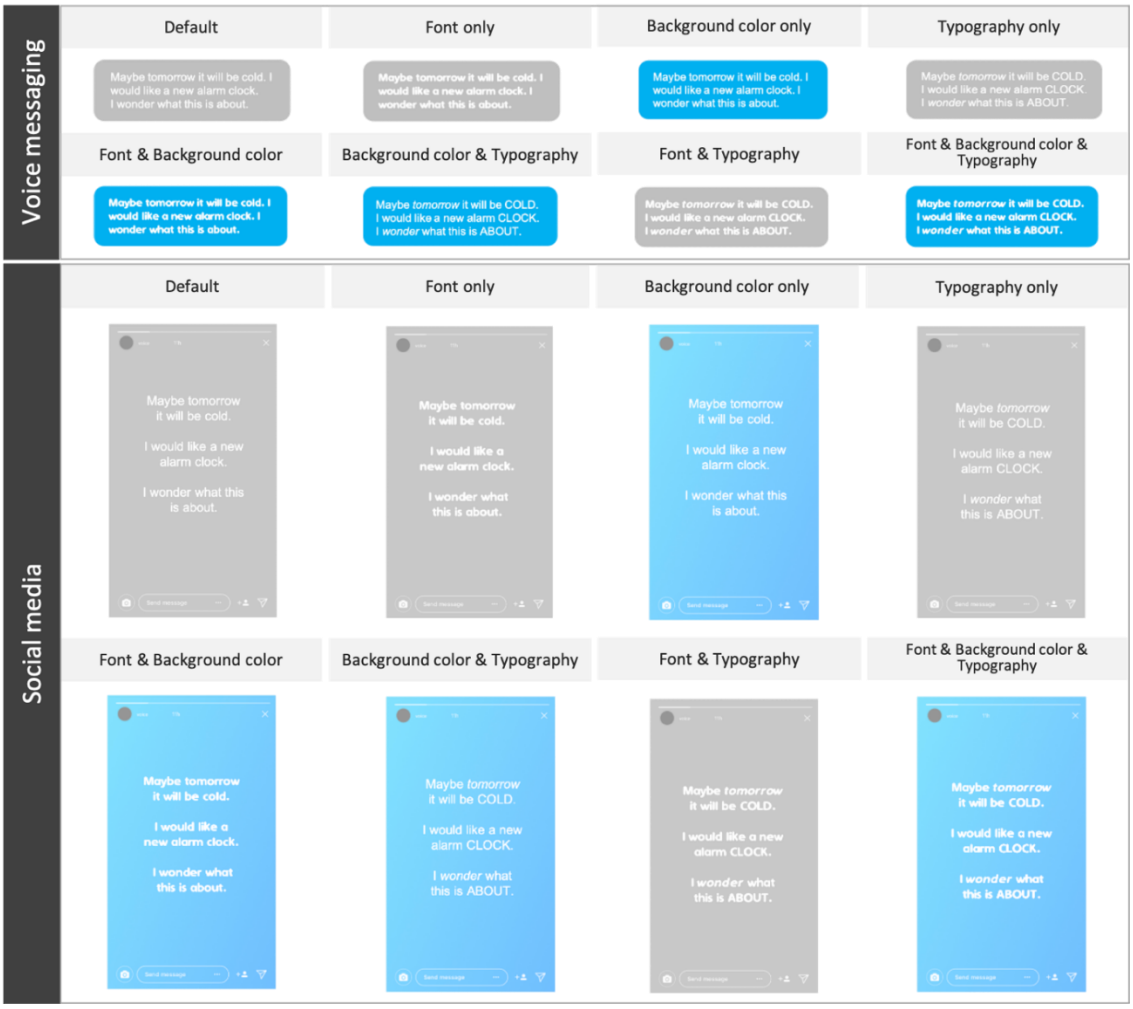

Figure 1: Examples of stimuli for "Happy" in voice messaging and social media.

Combining the font, background color and typography, we created seven text designs for each non-neutral emotion. Additionally, the text designs were integrated into two representative platforms: voice messaging and social media. To simulate the practical application, we applied the solid background color on the chat bubble for voice messaging. While for social media, we utilized the corresponding color as the gradient background. A total of 84 (7 text designs × 6 non-neutral emotions × 2 platforms) stimuli were created. Examples of stimuli for "Happy" are shown in Figure 1.

Experimental setup

A total of 31 participants voluntarily participated in the study (13 males and 18 females, ages ranging from 18 to 43, M = 26.23, SD = 5.62) and all were paid volunteers. The study was carried out in a web-based survey and the participants were given 84 stimuli in random order. As shown in Figure 2, for each voice stimuli, the participants were required to listen to the audio first, and compare the text designs with the default one based on three criteria: emotion expressivity, content delivery, and appropriateness, with a bipolar scale where -2 referred to "Strongly disagree" and +2 referred to "Strongly agree".

Figure 2: A screenshot of the web-based survey for this study: voice messaging and social media.

Results

Effect of font, background color and typography

We performed a three-way ANOVA to identify whether font background color and typography have an influence on emotion expressivity, content delivery, and appropriateness respectively.

Emotion expressivity

The results indicated that background color and typography statistically improved the emotion expressivity (background color: $F(1,2597) = 48.20$, $p < 0.05$; typography: $F(1,2597) = 67.92$, $p < 0.05$). Despite our participants' agreement that the font can express voice emotions to some extent ($M = 0.17 > 0$), there was no statistical difference observed in font ($F(1,2597) = 0.26$, $p > 0.05$). That is, our participants perceived a very limited affective impact of the font on the converted text in this experiment. We did not find any further statistically significant interaction effects.

Content delivery

There were statistically significant main effects of all the text designs (font: $F(1,2597) = 168.94$, $P < 0.05$; background color: $F(1,2597) = 49.13$, $p < 0.05$; typography: $F(1,2597) = 29.62$, $p < 0.05$). More

specifically, the users evaluated that background color (M = 0.34, SD = 1.04) and typography (M = 0.31, SD = 1.09) improved the content delivery of converted text, however the font even lowered the content delivery compared to the default (M = -0.10, SD = 1.09). Similarly, no significant interaction effects were found.

Appropriateness

As expected, font, background color and typography all had significant main effects on appropriateness (font: $F(1,2597) = 131.88$, $p < 0.05$; background color: $F(1,2597) = 93.56$, $p < 0.05$; typography: $F(1,2597) = 23.94$, $p < 0.05$). No statistically significant interaction effect was found. We see that background color received higher ratings (M = 0.21, SD = 1.02) than typography in appropriateness (M = 0.11, SD = 1.03). On the contrary, participants generally argued that using fonts was inappropriate (M = -0.20, SD = 0.04).

The most appropriate text design on voice messaging and social media

A one-way ANOVA was conducted to figure out which text design improved most in terms of emotion expressivity, content delivery, and appropriateness for two platforms. Figure 3 shows the mean rating scores that the different text design combinations received in two different platforms.

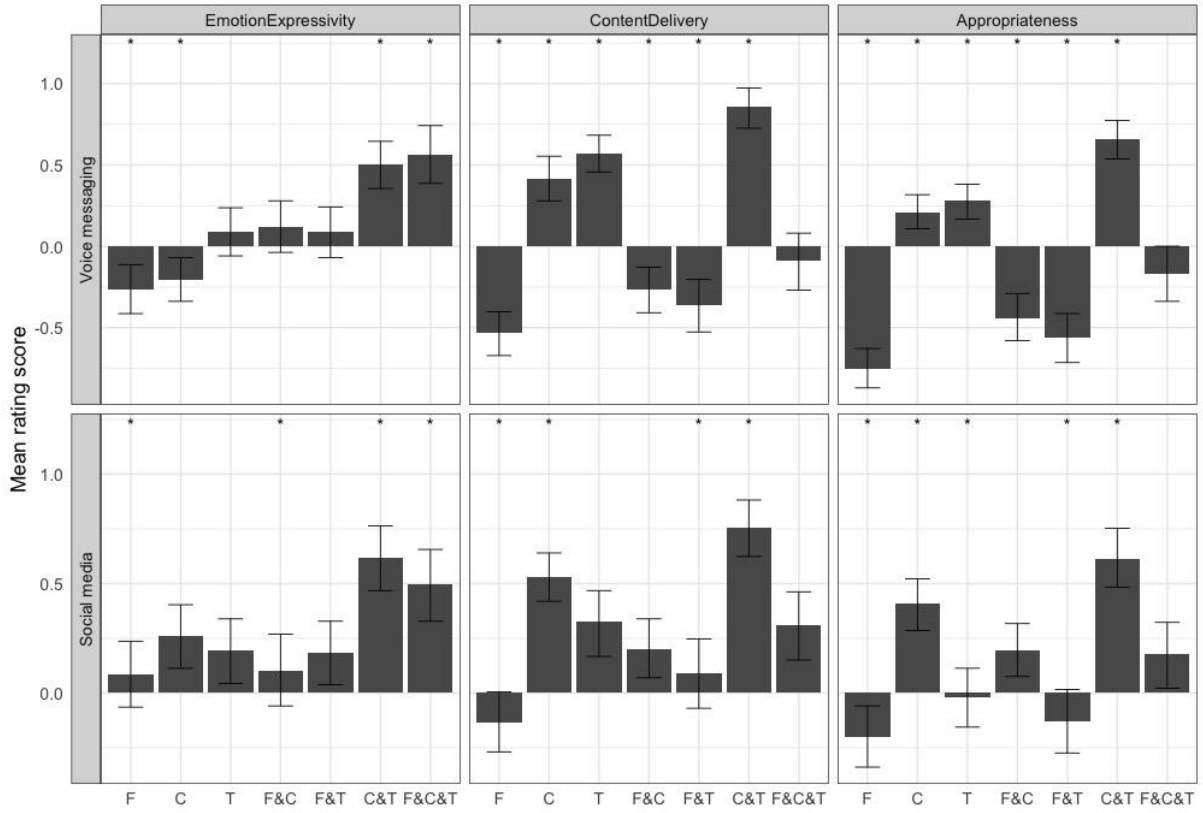

Figure 3: Average scales of emotion expressivity, content delivery and appropriateness in voice messaging and social media (F = "Font", C = "Background color", T = "Typography"). An asterisk indicates significance at p < .05 compared against all other text designs.

For voice messaging, "Background color & Typography" statistically received positive results in all three scales (emotion expressivity: M = 0.50, SD = 1.01; content delivery: M = 0.85, SD = 0.85; appropriateness: M = 0.66, SD = 0.87). This implies that the combination of background color and

typography not only improved the emotion expressivity and content delivery of converted text, but was also suitable for voice messaging. Surprisingly, although receiving the highest rating score in emotion expressivity (M= 0.56, SD = 1.20), "Font & Background color & Typography" was seen as negatively affecting content delivery and not suited in voice messaging. We also noticed the font had a negative effect on all three scales, and all text designs that used font (i.e., "Font & Background color": M = -0.45, SD = 1.02) were regarded inappropriate for use in voice messaging.

For social media, "Background color & Typography" received highest rating scores across all three scales (emotion expressivity: M = 0.62, SD = 1.04; content delivery: M = 0.75, SD = 0.92; appropriateness: M = 0.61, SD = 0.91). In terms of content delivery and appropriateness, the results show that "Background color" was also preferred, just below "Background color & Typography". Also, it was interesting to note that, different with voice messaging, participants agreed that "Font & Background color & Typography" significantly improved the three scales in social media. Nonetheless, despite statistical significance, the impacts of "Font & Background color & Typography" on content delivery and appropriateness are modest when compared to "Background color & Typography" and "Background color".

Conclusions

The effect of different text design elements was explored in this study. We found that, when compared to background color and typography, the effect of the font on emotion expressivity and content delivery of the converted text was insignificant and negative. The font is also seen as unsuitable for text design in two platforms, particularly voice messaging. The combination of background color and typography is shown to be a good text design to represent voice media in both voice messaging and social media. Although this study did not investigate all the variations of font, background color, and typography designs, it did demonstrate the potential of voice-aware text in narrowing the gap between voice and converted text, which can be used in affective voice interfaces and automatic captioning. For the future work, the influence of specific color and typography properties, such as color brightness and letter roundness need to be investigated.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF-2018R1A1A3A04078934) and the BK21 plus program through the National Research Foundation (NRF) funded by the Ministry of Education of Korea (NO.4120200913638).

References

1. de Lacerda Pataca C and Costa PD (2020), Speech modulated typography: towards an affective representation model, *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 139-143.
2. Choi S, Toshihiko Y and Kiyoharu A (2016), Typeface emotion analysis for communication on mobile messengers, *Proceedings of the 1st International Workshop on Multimedia Alternate Realities*, 37-40.
3. Chen Q, Yan Y and Suk H-J (2021), Bubble coloring to visualize the speech emotion, *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-6.

4. Hu J, Xu Q, Fu LP and Xu Y (2019), Emojilization: an automated method for speech to emoji-labeled text, *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1-6.
5. Cao H, Cooper DG, Keutmann MK, Gur RC, Nenkova A and Verma R (2014), Crema-d: Crowd-sourced emotional multimodal actors dataset, *IEEE Transactions on Affective Computing*, **5** (4), 377-390.
6. Hanada M (2018), Correspondence analysis of color–emotion associations, *Color Research and Application*, **43** (2), 224-237.
7. Johnson DO, Cuijpers RH and van der Pol D (2013), Imitating human emotions with artificial facial expressions, *International Journal of Social Robotics,* **5** (4), 503-513.
8. Suk H J and Irtel H (2010), Emotional response to color across media, *Color Research and Application,* **35** (1), 64-77.