Effects of Age on Live Streaming Viewer Engagement: A Dual Coding Perspective

Fei Liu

Department of Information, Technology and Innovation, Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China
Email: feiliu.sjtu@sjtu.edu.cn

Yijing Li

Department of Management and Marketing, The Hong Kong Polytechnic University, Hong Kong, China
Email: yijing1.li@polyu.edu.hk

Xiaofei Song

Department of Entrepreneurship, Marketing and Management Systems, University of Nottingham Ningbo China, Ningbo, China

Zhao Cai

Department of Entrepreneurship, Marketing and Management Systems, University of Nottingham Ningbo China, Ningbo, China
Email: zhao.cai@nottingham.edu.cn

Eric T.K. Lim

School of Information Systems and Technology Management, UNSW, Sydney, Australia
Email: e.t.lim@unsw.edu.au

Chee-Wee Tan

Department of Digitalization, Copenhagen Business School, Copenhagen, Denmark
Email: ct.digi@cbs.dk

Effects of Age on Live Streaming Viewer Engagement: A Dual Coding Perspective

Though the emerging live streaming industry has attracted growing attention, the prosperous yanzhi category where streamers mostly present daily chatting and amateur talent has not been widely investigated. To decode the mechanism behind the popularity of yanzhi category, this study posits that age estimated from a streamer's face and voice can influence the level of viewer engagement based on the proposition of dual coding theory. To test our hypothesized relationships, 274 one-minute video records ahead of a viewer commenting or/and gifting were collected and analyzed by deep learning algorithms. Analytical results attest to the negative association of both facial and vocal age with viewer engagement, while their interaction has a positive relationship with viewer engagement.

Keywords: live streaming; age; viewer engagement; dual coding theory; deep learning

Introduction

Since 2005, the live streaming industry in China has witnessed astonishing growth and is projected to attract around 524 million viewers by 2020 (Lin & Lu, 2017; Xinhuanet, 2019). The variability of live streaming content ranging from gaming to talent performance has amplified along with the growth of the industry. Accordingly, the *attraction economy* has experienced a dramatic rise that expedites the growth of derived sub-sectors like *yanzhi* (attractiveness index) live streaming that monetizes the public pursuing of a beautiful appearance (chinanews.com, 2019). In *yanzhi* category, streamers mainly interact with the audience via chatting and simple performance such as amateur singing and dancing. Although *yanzhi* streamers presented relatively low authenticity and proficiency, they obtained more attention and revenue than other categories such as outdoor and technology, according to a comparison of the popularity

index presented at the top of each sub-category in DouYu[1]. Instead of content quality, the exceptional importance of a streamer's physical characteristics such as age and attractiveness in *yanzhi* category facilitates an attempt to explicate how the individual perception like age estimation influences viewers' engagement actions in the likes of commenting and gifting. Streamers face a growing intensity of competition with over 3.5 million rivals in the live streaming industry (Cunningham et al., 2019). An investigation should be conducted to shed light on how streamers in this category could gain a rather high level of viewer engagement with relatively less informative content, to provide insights concerning the impact of age.

To disclose this phenomenon, the importance of viewer's initial impression on a *yanzhi* streamer should be emphasized, for it can influence their willingness to continue watching and engaging (Chen & Lin, 2018). Age has been identified as an important factor in shaping viewers' initial impression online, which is reflected by behaviors in dating apps such as fabricating profile information with a younger photo (Hall et al., 2010) or a lower age (Ellison et al., 2006). In the context of *yanzhi* live streaming, it implies that a streamer's age detected through his/her face and voice (Amilon et al., 2007; Moyse, 2014) can affect viewers' impressions. Given that, we posit that possessing a young appearance or voice contributes to a favorable impression through real-time communication and consequently transfer a passing-by viewer to a follower. Although a recent study has made preliminary exploration on the role of visual and vocal cues in the context of live streaming shopping (Sun et al., 2019), it is imperative to further scrutinize the mechanism by which viewer engagement is influenced by age estimated from visual and vocal cues in the context of *yanzhi* live streaming. The

[1] https://www.douyu.com/g_yz

current study is an endeavor to address the research gap by investigating the influence of visual and vocal age estimation of streamers on viewer engagement represented by sending comments and virtual gifts in *yanzhi* live streaming.

Regarding that, dual coding theory (DCT) is employed in this study to investigate the impact of age estimated from visual and vocal cues on viewer engagement. DCT elucidates that human cognition system consists of two distinct but inter-connected subsystems: a verbal one for processing linguistic information and a nonverbal one for processing symbolic characteristics of an object or event (Paivio, 2014). The incorporation of these two coding processes can mutually enhance the reliability of sensory estimations of age by integrating visual and vocal information (Belin et al., 2012). Since that, we posit that age estimated from verbal stimuli (voice of a streamer) and age estimated from non-verbal stimuli (facial appearance of a streamer) can be independently as well as interactively associated with viewer engagement.

To test the proposed hypotheses, we collected data from *yanzhi* category of a leading live streaming platform in China. The dataset contains 274 observations of viewer engagement actions. The age of a streamer was estimated from visual and vocal cues using a compact Soft Stagewise Regression Network (SSR-Net) algorithms and USTC Xunfei API, respectively. The ordered logistic regression analysis was employed to test the relationships between facial and vocal age and viewer engagement. This preliminary research attempts to make three theoretical contributions. First, this study extends live streaming literature by exploring the emerging *yanzhi* category where viewer engagement is determined by streamers' appearance and impression rather than the quality of content delivered by them. Second, DCT is incorporated to interpret the role of visual and vocal cues, which provides a novel insight to elucidate the mechanism of improving viewer engagement in live streaming. Third, this study validates the direct

and interaction effects of face and voice age on viewer engagement, enriching the live streaming literature with an emphasis on the imperative role of age estimated from visual and vocal cues, especially in categories with a less focus on content quality.

Theoretical Development and Hypotheses Formulation

Viewer Engagement in Yanzhi Category

The live streaming industry in China has grown rapidly with an increasing number of emerged online live streaming platforms such as HuYa, YY.com, and DouYu (Zheng & Jiang, 2016). Its market value was estimated from 5 billion US dollars in 2016 to 19 billion US dollars by 2022 (Cunningham et al., 2019). The competition among streamers has also intensified with the market size, resulting to a significant income gap: it was reported that 5% of 3.5 million professional streamers earned over 1,500 dollars per month, whereas the remaining gained less than 15 dollars per month in 2017 (Cunningham et al., 2019). The increasing popularity of full-time or part-time streamers facilitated the diversity of content categories including education, talent, daily life, video games, and e-commerce (Hu et al., 2017; Yu et al., 2018). Among emerging live streaming sub-categories, *yanzhi* category as a derivative of attraction economy phenomena has been paid insufficient attention in extant literature. Currently, the gaming category (Hilvert-Bruce et al., 2018; Recktenwald, 2017) and e-commerce category (Sun et al., 2019; Wongkitrungrueng & Assarut, 2020) have drawn the most attention in the live streaming domain.

In live streaming, streamers broadcast live content including gameplay, interactive chats, and talent show, meanwhile, viewers could engage through various approaches (e.g., commenting and sending virtual gifts) (Lin et al., 2021). Researchers incorporated the concept of engagement from marketing literature to study the

motivation of people participating in interactive activities such as gifting and commenting in the live streaming setting. *Viewer engagement* refers to the viewer's actions conducted to meet certain psychological needs such as entertainment and information seeking (Hilvert-Bruce et al., 2018).

Past research mostly paid attention to investigating the intrinsic motivations of viewer engagement. One stream of scholars depicted the live streaming viewer experience as a fulfilment process of individual needs (e.g., entertainment and social interaction). Hence, their studies are built on theories such as the uses and gratifications theory (UGT) (Hilvert-Bruce et al., 2018) and social identity theory (Hu et al., 2017). For instance, Hilvert-Bruce et al. (2018) proposed six intrinsic motivations for viewers engaging in the gaming category (i.e., social interaction, sense of community, meeting new people, entertainment, information seeking, and external support) building one UGT. Another stream treated it as a real-time multimedia communication process in which viewers expose to and interact with new technical features. Thus, they drew theories such as the unified theory of acceptance and use of technology (Sun et al., 2019) and IT affordance theory (Zheng & Jiang, 2016). Recently, growing attention has been drawn to the extrinsic motivations of viewer engagement. A recent empirical study found that the positive emotion of a streamer can extrinsically stimulate viewer engagement during a live streaming session (Lin et al., 2021). Apart from emotion, streamer information conveyed in the video content remains underexplored. Among emerging live streaming sub-categories, *yanzhi* category as a derivative of attraction economy phenomena has been paid insufficient attention in extant literature. Live streams in the gaming or e-commerce category mostly rely on improving content quality including an accentuated display of live gameplay and imparting relevant knowledge. Instead, the display of a streamer oneself combined with a profusion of

daily conversations and a lack of informative content highlight the importance of viewers' initial impression in *yanzhi* category. Consequently, it offers an opportunity to specifically scrutinize how viewer engagement is shaped by streamer appearance.

Different constructs including viewer's emotional connectedness, watching frequency, time spent, subscription time, continuance watching intention, and virtual gift have been used to operationalize this concept in live streaming (Hilvert-Bruce et al., 2018; Lu et al., 2018; Zheng & Jiang, 2016). Among these measures, gifting has drawn the most attention since it involves momentary rewards to a streamer and is directly related to their incomes (Yu et al., 2018). Recently, empirical research conducted by Zhou et al. (2019) showed that the action of sending comments is positively related to the virtual gifting decision, which implied that commenting can be regarded as a lower level of viewer engagement without monetary investment. Given that, the level of view engagement can be categorized by commenting only (low), gifting only (medium), and commenting with gifts (high).

To elucidate viewer behavior in the context of live streaming, literature highlighted the importance of investigating *viewer engagement* which refers to view's choice or action to meet certain psychological needs such as entertainment and information seeking (Hilvert-Bruce et al., 2018). Different constructs including viewer's emotional connectedness, watching frequency, time spent, subscription time, continuance watching intention, and virtual-gift have been used to operationalize this concept in live streaming (Hilvert-Bruce et al., 2018; Lu et al., 2018; Zheng & Jiang, 2016). Among these measures, gifting has drawn the most attention since it involves momentary rewards to a streamer and directly related to their incomes (Yu et al., 2018). Recently, an empirical research conducted by Zhou et al. (2019) showed that the action of sending comments is positively related to the virtual gifting decision, which implied

that commenting can be regarded as a lower level of viewer engagement without monetary investment. Given that, the level of view engagement can be categorized by commenting only (low), gifting only (medium), and commenting with gifts (high).

Age and Dual Coding Theory

In light of scrutinizing factors affecting viewer engagement, one stream of scholars depicted the live streaming viewer experience as a fulfilment process of individual-needs (e.g. entertainment and social interaction). Hence, their studies built on theories such as uses and gratifications theory (Hilvert-Bruce et al., 2018) and social identity theory (Hu et al., 2017). Another stream treated it as a real-time multimedia communication process in which viewers expose to and interact with new technical features. Thus, they drew theories such as unified theory of acceptance and use of technology (Sun et al., 2019) and IT affordance theory (Zheng & Jiang, 2016). In other words, existing studies mainly focused on content-quality-related attributes such as information seeking (Hilvert-Bruce et al., 2018) and guidance shopping affordance (Sun et al., 2019). Nonetheless, the characteristics of *yanzhi* category magnifies the priority of the impact of age formed through intrinsic even subconscious perceptions on a viewer's initial impression when studying viewer engagement. Combining with DCT, we postulate age estimated from visual and vocal stimuli are independently and interactively associated with viewer engagement, to fill the research gap in exploring *yanzhi* category.

The characteristics of *yanzhi* category highlight the key role of perceived age in affecting a viewer's initial impression of a streamer. Existing literature attested that people on the internet attempt to influence others' impressions by embellishing their appearance features that are immutable in face-to-face communication (Goffman, 2021; Moyse, 2014). Among these features, age is a vital one that can be assessed within few

seconds via observing the interlocutor, for an individual tend to react accordingly (Gladwell, 2006). The age reported in the profile acts as a filtering criterion when searching for potential dates in a dating app (Ward, 2017). Specifically, it has been shown in past research that age manipulation serves as the most common approach of viewers to manage impressions on their profiles, in online dating (Ellison et al., 2006; Ward, 2017). A study conducted by Ellison et al. (2006) has also revealed that viewers report one or two years younger to stay in a popular age range and survive the profile filtering. By observing the monthly rank of gifts received by streams in *yanzhi* category, most top-ranked streamers are relatively young with age-matching voices[2]. Likewise, live streaming viewers can decide whether to continue engaging based on their initial impression, which can be altered by the streamer's age.

DCT offers an appropriate lens to interpret how perceived streamer age impacts viewer engagement: visual and vocal cues are predominant sensory stimuli embodied in this video-mediated communication. According to the theory, human cognition develops two dependent but interrelated dual coding systems: a verbal channel for handling linguistic information and a nonverbal channel for processing nonverbal objects and events (Paivio, 2014). For live streaming viewers, the voice of a streamer is coded through their verbal subsystem, while the face of a streamer is handled via the nonverbal one. Despite the functional distinction between the two subsystems, cognitive coding process like the formation of age estimation also involves cross-subsystem activation of mental representations (Paivio, 2014). It means that verbal stimuli can trigger the nonverbal encoding process, meanwhile, verbal information can activate the nonverbal subsystem as well. Combining with DCT, we postulate age estimated from

[2] https://www.douyu.com/rank/rankList/tlist?cid=201

visual and vocal stimuli are independently and interactively associated with viewer engagement, to fill the research gap in exploring *yanzhi* category. We thus hypothesize that:

Hypothesis 1: *The age estimated from a streamer's face is negatively associated with the level of viewer engagement.*

Hypothesis 2: *The age estimated from a streamer's voice is negatively associated with the level of viewer engagement.*

Regarding the generation of age estimation, empirical evidence has affirmed that both face and voice could contribute to it exclusively (Moyse, 2014). Similarly, research in person perception supports that facial and vocal information containing idiosyncratic features like age and gender can shape individual perception during communication (Belin et al., 2012). With support from DCT, visual and vocal stimuli like age perceived from face and voice are coded via viewers' two cognitive subsystems interactively. Empirical evidence in individual perception further attests that the human brain integrates the sound of speech and the view of the speaker's face, for its interaction effects contribute to an enhanced information richness with more reliability during perception construction (Belin et al., 2012). It implies facial age estimation and vocal age estimation could interact to enforce their impact on viewer engagement. In other words, through the procedure of incorporating and crosschecking multi-channel age-related details, the congruence of perceived facial and vocal age is likely to enhance their influence on the level of viewer engagement. Since that, we postulate that:

Hypothesis 3: *The interaction between the age estimated from a streamer's face and the age estimated from the streamer's voice is positively associated with the level of viewer engagement.*

Methodology

Data were collected from the *yanzhi* category of DouYu which is a leading live streaming platform in China. Specifically, we extracted a one-minute cut from steamers' video streaming records ahead of their viewers' each observed engagement action. Viewer engagement actions (N=274) were transcribed into ordered three levels: 1 for commenting only, 2 for virtual gifting only, 3 for commenting together with virtual gifting, considering the amount of behavioural and monetary investment of viewers. The observed 127 streamers were selected on the front page of *yanzhi* category during random time slots from 8th to 22nd April 2019.

We adopted a deep learning approach to estimate each streamer's facial age and vocal age from her video clips with a Convolutional Neural Network (CNN) model, for it has been extensively applied to estimate appearance age based on identified facial features (Atallah et al., 2018). In a way, the deep learning approach can imitate the cognitive process through which human brains estimate a person's age based on appearance. Compared with the survey approach widely adopted in past live streaming research, the deep learning approach can enhance consistency and scalability while minimizing the biases caused by individual preferences. In particular, we adopt a novel CNN model named the Soft Stagewise Regression Network (SSR-Net) model designed by Yang et al. (2018) for age estimation. SSR-Net supports multi-class classification so that multiple age groups from 0 to above 80 with a middle point at 25 can be assigned for prediction. Moreover, multi-class classification can be performed in each stage of the multi-stage process so that each stage can improve the accuracy of age estimation iteratively. Apart from the multi-stage architecture, SSR-Net also supports a dynamic age range for every age class based on the input face image. Comparing to conventional CNN models that are large in size due to their complex neural networks, the SSR-Net

model is more efficient in processing the long image sequence in each video clip due to its compact size.

Likewise, the vocal age estimation of a video clip was performed using University of Science and Technology of China (USTC) Xunfei API[3], which interfaces the state-of-art deep-learning model for voice recognition. This API takes the audio sequence in each video clip as an input and estimates the vocal age three age groups: the young (below 12 years old), the middle-aged (12-40 years old), and the elderly (over 40 years old). Since the estimated facial age and vocal age heavily stew towards the younger side of the spectrum, we clustered the age groups of both facial age and vocal age either below or above the median. We then obtained two binary variables (i.e., *face age estimated* and *voice age estimated*) to indicate the younger and older groups for facial age and vocal age respectively (see in Table 1). Having these two binary variables also allows us to better evaluate the interaction effect between facial age and vocal age. As shown in Table 1, we controlled for variables that can potentially confound the effects of visual age and vocal age, including face beautifulness, gender (face and voice), emotion (face and voice), speech emotion, speed, and background music genre in the live streaming environment.

Though *viewer engagement* is a discrete dependent variable, multinomial logit regression is inappropriate, for it neglects the ordered nature of three levels of engagement efforts (Hanushek & Jackson, 2013). For this reason, we applied the ordered logit regression method using the *ologit* package in STATA/SE 15.1 for hypothesis testing. *Viewer engagement* is operationalized as ordinal under the assumption that the levels of viewer engagement are naturally ordered categories from

[3] https://www.xfyun.cn/services/sound-feature-recg

commenting only to both commenting and gifting. We configured Model 1 to test the direct effects of *face age estimated* and *voice age estimated* on *viewer engagement*. We then built Model 2 to test the interaction effect denoted by *face age estimated * voice age estimated*.

Analytical Results

Table 2 summarizes the estimation results of the ordered logistic regression, which support all three hypotheses. The effect size of each path coefficient is determined by the odds ratio calculated by exponentiating the corresponding path coefficient. The estimating result of Model 1 shows that both *face age estimated* and *voice age estimated* impose a significant negative impact on the level of viewer engagement. While holding others variable constant, if a streamer's facial age was perceived as above 25 years old as opposed to equal to or below 25 years old, the log odds of achieving a higher viewer engagement level would be reduced by 1.971. Likewise, with other variables held constant, when a streamer's vocal age was perceived as perceived as above 12 years old as opposed to equal to or below 12 years old, the log odds of achieving a higher viewer engagement level was expected to decrease by 0.624. Again, these results confirmed the negative effects of facial age and vocal age on viewer engagement as posit by *Hypothesis 1 and 2*.

Model 2 reported a higher Pseudo R^2 with a significantly large Likelihood-ratio test statistic. This result shows that by adding the interaction term in Model 2, it can better explain the variance in the dataset comparing to Model 1. The interaction effect between facial age and vocal age in Model 2 exerts a significant positive impact on the log odds of viewer engagement level. Specifically, with other variables held constant, when a streamer with a more mature face that is matched with a more mature voice, the log odds of achieving a higher viewer engagement level would be increased by 1.359.

Upon a more in-depth analysis, as shown in Figure 1, the impact of vocal age on the probability of achieving a desirable viewer engagement level is much more substantial for a streamer with a less mature face. Taken together, the result helps to substantiate the interaction effect between the facial age and the vocal age on viewer engagement as posited in *Hypothesis 3*.

Discussion and Implication

This study inspects the role of age estimation generated from visual and vocal stimuli via the dual coding systems played in influencing viewer engagement in the *yanzhi* live streaming context. It is hypothesized that age estimation from a streamer's face and voice is negatively associated with viewer engagement. In other words, younger age perceived from face or voice is related to a higher level of engagement toward commenting with gifting. A positive relationship between the interaction of two factors and viewer engagement is also postulated. With the application of deep learning algorithms, the age of a streamer was estimated based on facial or vocal features extracted from one-minute live streaming video records ahead of a viewer's engagement action (i.e. commenting or gifting). Based on the output, ordered logistic regression was utilized to test hypotheses. Estimation results provide empirical evidence for the impact of age estimation on viewer engagement through dual coding channels. Specifically, it is confirmed that facial age visually conveyed by a streamer's physical appearance negatively contributes to the viewer's engagement level, which means that a more mature face is associated with less investment in engaging live streaming activities such as commenting and gifting. Apart from that, this study also shows that a streamer's age as a latent message sent through his/her voice can positively contribute to a higher viewer engagement level. Both of them align with the finding in dating app context where age manipulation in profile serves as an impression management approach to

attract more potential dates (Ellison et al. 2006; Moyse 2014). Furthermore, this study attests to the positive impact of interaction between facial and vocal age estimation on viewer engagement implied by DCT (Paivio 2014) and individual perception literature (Belin et al. 2012). It denotes that a streamer possessing a young face and voice simultaneously can exploit most benefits concerning increasing viewer engagement level.

There are three contributions of this study made to the existing literature. First, it extends extant live streaming domain by exploring the emerging yanzhi category, which presents an opportunity for researchers to investigate viewer engagement behaviors with an emphasized importance of sensory factors influencing engaging in real-time communication between streamers and viewers. Gaining insights from this area is helpful to decompose the mechanism of improving viewer engagement starting from fundamental cognitive level. Second, this study complements prior research in live streaming viewer engagement that highlights one type of sensory information like vision or neglected their interaction effect (Park & Lin, 2020; Sun et al., 2019). Building on DCT, this research posits and confirms that perceived facial age and vocal age along with their interaction have a significant impact on viewer engagement. Third, this study also demonstrates that the integration of facial and vocal age estimation has an enforcement effect on contributing to encouraging a higher level of viewer engagement if a streamer's face and voice are perceived young with other factors unchanged. Since that, it not only empirically examines the effect of age estimation on viewer behavior to enrich the age estimation domain, also echoes the person perception studies in the integration of visual and vocal cues (Belin et al., 2012).

Limitations and Future Research

This study has two limitations that bring research potentials for the future as well. First,

the sample size of this study is sufficient for a preliminary test but needs to be enlarged to incorporate more streamers and user engagement observations. By doing so, a solid estimation on the impact of personal characteristics and behavioral differences can be generated. Second, more control variables could be integrated into the regression model, as DCT alludes that transcribed verbal cues including and nonverbal stimuli such as voice sentiment and body language can also be coded via the cognitive dual system and consequently shape human intention and behaviors (Paivio 2014).

References:

Amilon, K., Weijer, J. v. d., & Schötz, S. (2007). The impact of visual and auditory cues in age estimation. In *Speaker Classification II* (pp. 10-21). Springer.

Atallah, R. R., Kamsin, A., Ismail, M. A., Abdelrahman, S. A., & Zerdoumi, S. (2018). Face recognition and age estimation implications of changes in facial features: A critical review study. *IEEE Access*, *6*, 28290-28304.

Belin, P., Campanella, S., & Ethofer, T. (2012). *Integrating face and voice in person perception*. Springer Science & Business Media.

Chen, C.-C., & Lin, Y.-C. (2018). What drives live-stream usage intention? The perspectives of flow, entertainment, social interaction, and endorsement. *Telematics and Informatics*, *35*(1), 293-303.

chinanews.com. (2019). *Consumption Upgrade Brings "Yanzhi Jingji", Half of Lipstick Purchasers Are Men*.

Cunningham, S., Craig, D., & Lv, J. (2019). China's livestreaming industry: platforms, politics, and precarity. *International Journal of Cultural Studies*, *22*(6), 719-736.

Ellison, N., Heino, R., & Gibbs, J. (2006). Managing impressions online: Self-presentation processes in the online dating environment. *Journal of Computer-Mediated Communication*, *11*(2), 415-441.

Gladwell, M. (2006). Blink: The power of thinking without thinking.

Goffman, E. (2021). *The presentation of self in everyday life*. Anchor.

Hall, J. A., Park, N., Song, H., & Cody, M. J. (2010). Strategic misrepresentation in online dating: The effects of gender, self-monitoring, and personality traits. *Journal of Social and Personal Relationships*, *27*(1), 117-135.

Hanushek, E. A., & Jackson, J. E. (2013). *Statistical methods for social scientists*. Academic Press.

Hilvert-Bruce, Z., Neill, J. T., Sjöblom, M., & Hamari, J. (2018). Social motivations of live-streaming viewer engagement on Twitch. *Computers in Human Behavior*, *84*, 58-67.

Hu, M., Zhang, M., & Wang, Y. (2017). Why do audiences choose to keep watching on live video streaming platforms? An explanation of dual identification framework. *Computers in Human Behavior*, *75*, 594-606.

Lin, J., & Lu, Z. (2017). The rise and proliferation of live-streaming in China: Insights and lessons. International conference on human-computer interaction,

Lin, Y., Yao, D., & Chen, X. (2021). Happiness Begets Money: Emotion and Engagement in Live Streaming. *Journal of Marketing Research*, *58*(3), 417-438. https://doi.org/10.1177/00222437211002477

Lu, Z., Xia, H., Heo, S., & Wigdor, D. (2018). You watch, you give, and you engage: a study of live streaming practices in China. Proceedings of the 2018 CHI conference on human factors in computing systems,

Moyse, E. (2014). Age estimation from faces and voices: A review. *Psychologica Belgica*, *54*(3).

Paivio, A. (2014). *Mind and its evolution: A dual coding theoretical approach*. Psychology Press.

Park, H. J., & Lin, L. M. (2020). The effects of match-ups on the consumer attitudes toward internet celebrities and their live streaming contents in the context of product endorsement. *Journal of Retailing and Consumer Services*, *52*, 101934.

Recktenwald, D. (2017). Toward a transcription and analysis of live streaming on Twitch. *Journal of Pragmatics*, *115*, 68-81.

Sun, Y., Shao, X., Li, X., Guo, Y., & Nie, K. (2019). How live streaming influences purchase intentions in social commerce: An IT affordance perspective. *Electronic Commerce Research and Applications*, *37*, 100886.

Ward, J. (2017). What are you doing on Tinder? Impression management on a matchmaking mobile app. *Information, Communication & Society*, *20*(11), 1644-1659.

Wongkitrungrueng, A., & Assarut, N. (2020). The role of live streaming in building consumer trust and engagement with social commerce sellers. *Journal of Business Research*, *117*, 543-556.

Xinhuanet. (2019). *Competition in China's Live-Streaming Market Gets Fiercer: Report*. http://www.xinhuanet.com/english/2019-09/11/c_138384107.htm

Yang, T.-Y., Huang, Y.-H., Lin, Y.-Y., Hsiu, P.-C., & Chuang, Y.-Y. (2018). Ssr-net: A compact soft stagewise regression network for age estimation. IJCAI,

Yu, E., Jung, C., Kim, H., & Jung, J. (2018). Impact of viewer engagement on gift-giving in live video streaming. *Telematics and Informatics*, *35*(5), 1450-1460.

Zheng, Y., & Jiang, C. (2016). Investigating the Impact Factors Forming Users' Intention in Utilizing Live Online Platforms: Case Study DouYu TV. 2016 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII),

Zhou, J., Zhou, J., Ding, Y., & Wang, H. (2019). The magic of danmaku: A social interaction perspective of gift sending on live streaming platforms. *Electronic Commerce Research and Applications*, *34*, 100815.

Table 1. Variable List and Descriptive Statistics

Variable	Coding Method	Measurement	Observation	Mean	Std.Dev	Min	Max
Viewer engagement	N/A	1: commenting only, 2: gifting only, 3: commenting with gifting	274	2.088	0.789	1	3
Face age estimated	SSR-Net	1: above 25 years old, 0: below 25 years old (include)	304	0.434	0.496	0	1
Voice age estimated	USTC Xunfei API[4]	1: above 12 years old, 0: below 12 years old (include)	304	0.520	0.500	0	1
Face age estimated * Voice age estimated	N/A	the interaction term	304	0.283	0.451	0	1
Face beautifulness	USTC Xunfei API	0: very beautiful, 1: beautiful	274	0.365	0.482	0	1
Face sex	SSR-Net, gender model	1: female, 0: male, 0.5: unisex	274	0.894	0.353	0	2
Voice female	Long Short Term Memory Neurtal Networks[5]	The probability of female	304	0.564	0.144	0.163	0.899
Z (Voice female)^2	N/A	Squared Z score of Voice female	304	0.997	1.488	0.000	7.688
Voice emotion	The Ryerson Audio-Visual Database of Emotional Speech and Song[6]	1: positive, 0: non-positive	304	0.766	0.347	0	1
Speech emotion	Pre-Trained Chinese BERT with Whole Word Masking[7]	1: positive, 0: non-positive	304	0.354	0.080	0.150	0.674
Speech speed	N/A	Number of words/second	304	7.851	3.704	0.631	24.926
Speech speed std	N/A	Standard deviation of speed	274	4.923	2.505	0.640	14.478
Music genre	Compact CNN[8]	1: pop, 0: other	304	0.806	0.396	0	1

[4] https://www.xfyun.cn/doc/voiceservice/sound-feature-recg/API.html

[5] https://github.com/JinScientist/voice-gender-recognition

[6] https://github.com/marcogdepinto/Emotion-Classification-Ravdess

[7] https://github.com/ymcui/Chinese-BERT-wwm

[8] https://github.com/keunwoochoi/music-auto_tagging-keras

Table 2. Ordered Logistic Regression Estimation Results

Variable	Model 1		Model 2	
	Coefficient	**Odds Ratio**	**Coefficient**	**Odds Ratio**
Independent Variables				
Face age estimated	-2.074***	0.126***	-2.822***	0.060***
Voice age estimated	-0.713***	0.490***	-1.277***	0.279***
Interaction Term				
Face age estimated*Voice age estimated			1.399***	4.050***
Control Variables				
Face sex	0.716	2.047	0.732	2.079
Voice female	1.258	3.520	1.165	3.205
Z (Voice female)^2	-0.209*	0.812*	-0.192	0.826
Voice emotion	0.330	1.391	0.450*	1.568*
Speech emotion	3.689*	39.987*	3.442*	31.237*
Speech speed	0.053	1.054	0.061	1.063
Speech speed std	0.212***	1.236***	0.203***	1.225***
Music genre	0.981*	2.666*	1.018*	2.768*
Observation	274	274	274	274
Pseudo R^2	0.2513	0.2513	0.2631	0.2631
Likelihood-ratio test χ^2	7.03***			

***Note*:** * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Figure 1. Interaction Effect of Face Age Estimated and Voice Age Estimated on Viewer Engagement