

Impact of Anchors' Voice Characteristics on Viewer Engagement in Live Streams

Qingli Zeng

Hebrew University of Jerusalem

Sandeep Chandukala

Singapore Management University

Hai Che

University of California at Riverside

Lifeng Yang

Shanghaitech University

Cite as:

Zeng Qingli, Chandukala Sandeep, Che Hai, Yang Lifeng (2026), Impact of Anchors' Voice Characteristics on Viewer Engagement in Live Streams. *Proceedings of the European Marketing Academy*, 55th, (133288)

Paper from the 55th Annual EMAC Conference, Bath, United Kingdom, June 2-5, 2026



Impact of Anchors' Voice Characteristics on Viewer Engagement in Live Streams

Abstract

In the rapidly growing live-streaming market, anchors primarily communicate vocally while viewers engage through both immediate active participation (chats) and sustained commitment (subscriptions). Drawing upon dual-process frameworks of social approval and social influence, this study examines how anchors' expressive energy, stressed vocal tones, and cognitive depth differentially affect these distinct engagement behaviors. Leveraging over 1,000 hours of gaming content from top streamers, we employ advanced audio, video, and text analytics to uncover nuanced engagement patterns. Results indicate that expressive energy and cognitive vocal expressions enhance sustained commitment without significantly influencing short-term chat participation, whereas stressed tones positively impact both engagement types. Furthermore, contextual moderators such as chatroom sentiment and channel popularity significantly alter these relationships, revealing that vocal cues affect viewer engagement differently depending on the social environment. This research contributes theoretically by delineating distinct psychological mechanisms underlying viewer engagement and offers actionable guidance for platforms and content creators.

Keywords: *audio analysis, live streaming, multimodal analysis*

Track: Digital Marketing & Social Media

1. Introduction

The live-streaming gaming market is expanding rapidly, with platforms like Twitch and YouTube Gaming redefining digital entertainment. In this highly competitive environment, understanding the mechanisms that drive viewer engagement is critical for platform success. Engagement in live streaming manifests in two primary forms: immediate interactive behaviors (e.g., real-time chats) and sustained commitment behaviors (e.g., subscriptions and financial support). While immediate behaviors signify momentary interest and community activity, sustained commitment reflects loyalty and generates reliable revenue.

Unlike traditional broadcasts or social influencers who rely heavily on facial expressions and gestures, gaming anchors face unique constraints. The anchor's video feed is typically minimized relative to the gameplay footage, and because their hands are occupied, they cannot communicate via text chat. Consequently, the anchor's voice serves as the critical medium of communication, delivering not only information but also emotional and cognitive cues that influence viewer perceptions.

Despite the importance of vocal delivery, prior research has largely examined engagement metrics in isolation or focused on prepared, one-way communications such as earnings calls or ads (Cabeza-Ramírez et al., 2022; Johnson & Woodcock, 2019). Existing studies on live streaming have explored factors like broadcaster happiness (Lin et al., 2021) or audience size (Lu et al., 2021) on tipping behavior, but there is limited knowledge about how specific vocal characteristics—such as affective and cognitive expressions—impact the distinct types of viewer engagement. Furthermore, the effects of these vocal cues likely depend on the social context, specifically the chatroom sentiment and channel popularity.

To address these gaps, this study integrates Social Approval Theory and Social Influence Theory. Social Approval Theory suggests that communicators who effectively employ emotional appeals generate positive evaluations and greater support. Conversely, Social Influence Theory posits that individual responses to persuasion are shaped by social contextual factors (Cialdini, 2006), such as the collective mood of the crowd (sentiment) or the perceived status of the source (popularity). We distinguish between three key dimensions of vocal expression: expressive energy (enthusiasm), stressed tones (tension/effort), and cognitive tones (analytical thinking).

We analyze a dataset comprising more than 1,000 hours of content from the top 25 anchors on a prominent U.S. live-streaming platform. By utilizing audio mining technology alongside video and text analytics, we capture intricate nonverbal cues while controlling for video dynamics and viewer sentiment. To account for the endogeneity of vocal characteristics, we employ an instrument-free copula approach (Park & Gupta, 2012).

Our research makes three key contributions. First, we demonstrate that different vocal characteristics influence distinct engagement types through different pathways; for instance, expressive energy drives subscriptions but not chats. Second, we show that social contextual factors moderate these effects; channel popularity amplifies the impact of expressive energy on subscriptions while dampening its effect on chats. Third, we introduce a methodological template for analyzing complex multimodal interactions in digital environments.

2. Theoretical Background

To understand how vocal characteristics drive engagement, we integrate **Social Approval Theory** and **Social Influence Theory**.

2.1 Vocal Characteristics and Social Approval

Social Approval Theory explains how audiences form evaluations about communicators based on emotional and intuitive processes (Bundy & Pfarrer, 2015). Communicators who

effectively employ emotional appeals generate more positive evaluations and greater support.

For instance, Berger and Milkman (2012) demonstrated that content characteristics like emotionality significantly alter social approval evaluations. In the context of live streaming, we focus on three distinct dimensions of vocal expression:

- **Expressive Energy:** Conveying enthusiasm and emotional intensity.
- **Stressed Tones:** Indicating tension, pressure, or effort.
- **Cognitive Tones:** Reflecting analytical thinking and information processing.

We propose that anchors whose vocal characteristics convey positive energy and competence will receive higher social approval, enhancing sustained commitment (subscriptions). While immediate interactive behaviors (chats) may be impulsive, sustained commitment represents a deeper loyalty driven by these evaluations.

2.2 The Moderating Role of Social Context

Social Influence Theory (SIT) posits that individual responses to persuasive communication are shaped by the social environment. In live streaming, this context is defined by two key factors:

Chatroom Sentiment: The collective emotional tone of viewer interactions. Positive sentiment may satisfy viewers' social needs, potentially dampening the urgency to respond to anchor cues.

Channel Popularity: The size and status of the audience. Popularity acts as social proof; popular anchors may benefit from a "halo effect," granting them greater flexibility in vocal delivery, whereas emerging anchors may need to exert more vocal effort to achieve similar engagement.

We propose that while vocal characteristics influence engagement through social approval mechanisms, these effects are significantly moderated by the social context created by channel popularity and chatroom sentiment.

3. Data and Methodology

3.1 Data Overview

We collected data from a prominent U.S. gaming live-streaming platform, focusing on the top 25 anchors who account for 35% of the platform's viewership. This ensures we are analyzing successful communicators, though we acknowledge potential selection bias. The dataset comprises over 1,000 hours of content, including audio streams (40,000 Hz), high-definition video, chat logs, and subscription records.

3.2 Variables and Measurement

Dependent Variables: We operationalize engagement into two distinct count variables aligned with existing research standards (Lin et al., 2021; Lu et al., 2021):

Immediate Interactive Behaviors: The frequency of chats per minute.

Sustained Commitment Behaviors: The frequency of new subscriptions per minute.

Independent Variables (Vocal Characteristics): We utilized the Nemesysco QA5 audio analysis system to extract emotion-related features from the anchor's voice. This software has been validated in prior academic research (Hobson et al., 2012; Wang et al., 2021). Due to high correlations among raw audio features, we conducted a factor analysis (Eigenvalue > 1), which yielded three distinct factors accounting for 68% of the variance:

Expressive Energy: Loaded heavily by energy, content, and excitement variables.

Stressed: Loaded by stress indicators.

Cognition: Loaded by cognitive and analytical markers.

Moderators:

Chatroom Sentiment: Calculated using the NLTK package in Python to generate an average sentiment score for all chat messages within each minute (Bird et al., 2009) .

Channel Popularity: A binary variable indicating whether the channel was among the top five channels at the time of streaming, representing high visibility and status (Li et al., 2024).

Control Variables: We included an extensive set of controls to isolate the effects of vocal tones:

Video Dynamics: We adapted the Structural Similarity Index (SSIM) to measure visual changes. We calculated **Within-Minute Dynamics** (similarity between frames within a minute) and **Across-Minute Dynamics** (similarity between consecutive minute intervals).

Visual Attributes: Entropy (image detail), Hue, Saturation, and Value (brightness).

Speech Quality: MOSNet scores to control for audio clarity.

Lagged Metrics: Previous minute's chat volume and subscriptions to account for temporal dependence.

Anchor Fixed Effects: To control for unobserved heterogeneity.

3.3 Empirical Model and Endogeneity

Since our dependent variables are count data (chats and subscriptions), we employed a Negative Binomial Regression model. The estimation model is specified as:

$$\mu_{i,t} = \beta_{Audio} \cdot Audio_{i,t-1} + \beta_{Mod} \cdot Moderators_{i,t-1} + \beta_{Inter} \cdot (Audio \times Moderators) + \beta_{Controls} \cdot Controls_{i,t-1} + \lambda_i + \epsilon_{i,t}$$

Addressing Endogeneity: Recognizing that viewer behavior might influence an anchor's voice (reverse causality) or that unobserved confounders may exist, we employed the instrument-free Gaussian Copula approach. We generated copula correction terms for the three focal audio variables and included them in the regression to correct for potential endogeneity.

4. Empirical Results

4.1 Main Effects

Our baseline analysis (omitted for brevity) reveals that different vocal characteristics influence engagement through distinct pathways. We find that stressed vocal tones significantly increase both immediate interactive behaviors (chats) ($\beta = 0.0037, p < 0.01$) and sustained commitment (subscriptions) ($\beta = 0.0377, p < 0.01$). In contrast, expressive energy ($\beta = 0.0502, p < 0.01$) and cognitive tones ($\beta = 0.044373, p < 0.05$) significantly enhance sustained commitment but do not have a statistically significant direct effect on immediate interactive behaviors.

This divergence suggests that while stressed tones—potentially signaling authenticity or effort—trigger broad engagement, the positive evaluation required for financial commitment (subscriptions) is specifically driven by the anchor’s energy and cognitive depth.

4.2 Moderating Effects of Social Context

Table 1 presents the results of our primary model (Negative Binomial Regression), which includes interaction terms to test the moderating effects of channel popularity and chatroom sentiment.

Table 1: Main Model with Endogeneity Correction (Condensed)

Variable	Immediate Interactive Behaviors (Chats)	Sustained Commitment Behaviors (Subs)
Expressive Energy	0.0099***	0.0377***
× Channel Popularity	-0.0177***	0.0285***
× Chatroom Sentiment	0.0173	-0.302***
Stressed	0.0004	0.0403***
× Channel Popularity	0.0094***	-0.0226***
× Chatroom Sentiment	-0.0090**	-0.0051
Cognitive Level	0.0021	0.0386**
Channel Popularity	0.922***	-0.355***

Chatroom Sentiment	-0.0649	0.654***
--------------------	---------	----------

*** p<0.01, ** p<0.05, * p<0.1

Note: Coefficients for control variables including video dynamics, visual attributes, and lagged metrics are included in estimation but omitted here for brevity.

4.3 Interaction Analysis

The interaction results highlight the critical role of social context:

Popularity and Expressive Energy: Channel popularity amplifies the positive effect of expressive energy on sustained commitment ($\beta = 0.0285, p < 0.01$). This suggests that viewers of popular channels, having already validated the anchor's status, place greater weight on expressive delivery when making subscription decisions. However, popularity dampens the effect of energy on immediate chats ($\beta = -0.0177, p < 0.01$), implying that in large, active communities, the anchor's own energy is less critical for stimulating chat activity.

Sentiment and Expressive Energy: Positive chatroom sentiment negatively moderates the impact of expressive energy on subscriptions ($\beta = -0.302, p < 0.01$). When the social environment is already positive, the incremental value of the anchor's energy in driving subscriptions diminishes, consistent with Social Influence Theory.

Popularity and Stress: Stressed tones enhance chat activity more effectively in popular channels ($\beta = 0.0094, p < 0.01$), potentially acting as a rallying cry for the community. Conversely, the positive impact of stress on subscriptions is weaker for popular channels ($\beta = -0.0226, p < 0.01$), suggesting that while stress drives chatter, established audiences may find excessive tension less appealing for long-term financial commitment.

4.4 Robustness Checks

We employed the instrument-free copula approach to address endogeneity; the significance of the copula terms ($p < 0.05$) confirmed the necessity of this correction.

Furthermore, we re-estimated the model using a two-minute lag (t-2) for independent

variables. The results remained largely consistent, though the significance of cognitive levels attenuated, suggesting cognitive cues may have a more transient effect than emotional ones.

5. Discussion

5.1 Theoretical Implications

This study advances the literature on digital engagement by demonstrating that engagement is not a unidimensional construct. We show that immediate interactive behaviors and sustained commitment behaviors are driven by distinct psychological mechanisms. While Social Approval Theory explains why expressive and cognitive qualities drive subscriptions (positive evaluation), Social Influence Theory explains why these effects vary by context. Notably, we identify a "stress paradox": contrary to the intuition that stress is negative, stressed vocal tones positively impact engagement, likely functioning as a signal of authenticity and emotional investment.

5.2 Managerial Implications

For content creators and platforms, our findings offer strategic guidance for vocal delivery (Table 2).

Table 2: Strategic Recommendations

Goal	Strategy	Rationale
New Anchors	Match Sentiment: Align vocal energy with positive chat sentiment.	Positive sentiment amplifies the impact of energy on interaction for smaller audiences.
Established Anchors	Differentiate Tones: Use expressive energy to drive subscriptions but rely on popularity to drive chats.	Popularity naturally drives chats; energy is best reserved for converting viewers to subscribers.
Authenticity	Leverage Stress: Use stressed tones to signal effort but monitor intensity.	Stress drives engagement but can backfire on subscriptions if overused in popular channels.

5.3 Limitations and Future Research

Our study focused on the top 25 gaming anchors to ensure data quality, which may limit generalizability to smaller, niche streamers. Future research should explore a broader

range of anchors and content categories (e.g., "Just Chatting" or educational streams) to validate these findings across different contexts. Additionally, while we controlled for video dynamics, future work could investigate specific visual semantic features (e.g., facial expressions) in conjunction with vocal tones.

6. Conclusion

In the visually crowded environment of live streaming, the anchor's voice is a powerful tool for differentiation. By strategically modulating expressive energy, stress, and cognitive tones based on channel popularity and real-time sentiment, anchors can optimize both community activity and revenue generation. This research highlights that effective digital communication is not static but dynamically interacts with the social context of the audience.

Selected key references

- Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2), 192–205.
- Bundy, J., & Pfarrer, M. D. (2015). A burden of responsibility: The role of social approval at the onset of a crisis. *Academy of Management Review*, 40(3), 345–369.
- Chen, H., Dou, Y., & Xiao, Y. (2023). Understanding the role of live streamers in live-streaming e-commerce. *Electronic Commerce Research and Applications*, 59, 101266.
- Cialdini, R. (2006). *The psychology of persuasion* (Rev. ed.). Harper Business.
- Johnson, M. R., & Woodcock, J. (2019). The impacts of live streaming and Twitch.tv on the video game industry. *Media, Culture & Society*, 41(5), 670–688.
- Li, L., Chen, X., & Zhu, P. (2024). How do e-commerce anchors' characteristics influence consumers' impulse buying? An emotional contagion perspective. *Journal of Retailing and Consumer Services*, 76, 103687.
- Lin, Y., Yao, D., & Chen, X. (2021). Happiness begets money: Emotion and engagement in live streaming. *Journal of Marketing Research*, 58(3), 417–438.
- Liu, H., Tan, K. H., & Pawar, K. (2022). Predicting viewer gifting behavior in sports live streaming platforms: The impact of viewer perception and satisfaction. *Journal of Business Research*, 144, 599–613.
- Lu, S., Yao, D., Chen, X., & Grewal, R. (2021). Do larger audiences generate greater revenues under pay what you want? Evidence from a live streaming platform. *Marketing Science*, 40(5), 964–984.
- Park, S., & Gupta, S. (2012). Handling endogenous regressors by joint estimation using copulas. *Marketing Science*, 31(4), 567–586.
- Wang, X., Lu, S., Li, X. I., Khamitov, M., & Bendle, N. (2021). Audio mining: The role of vocal tone in persuasion. *Journal of Consumer Research*, 48(2), 189–211.