Journal of
Theoretical and Applied Electronic Commerce Research

Article

Beyond Words: How Streamers' Dynamic Nonverbal Cues Increase Consumer Purchase Behavior Through Viewer Immersion

Xiaochen Liu [1], Tianyang Ma [2,*], Qianqian Han [2] and Qiang Yang [2]

[1] School of Economics and Management, Liaoning University of Technology, Jinzhou 121010, China; claireliu2023@gmail.com
[2] School of Business, Nanjing Audit University, Nanjing 211815, China; ma2403026@stu.nau.edu.cn (Q.H.)
* Correspondence: mp2403013@stu.nau.edu.cn

Abstract

Live-streaming commerce has become a routine channel for merchants, and streamers' nonverbal cues are closely associated with consumer responses and conversion. Drawing on real live-streaming settings, this study examined the relationship between streamers' nonverbal cues and consumer purchase behavior, and further tested whether immersion, as reflected by average watch time, helped explain this relationship. Building on Social Cognitive Theory, we constructed a multimodal dataset of 4600 product-presentation segments from 546 live sessions. Using an automated computer-vision-based framework, we measured segment-level nonverbal behaviors, including nodding frequency, gesture intensity, postural movement intensity, forward lean, and camera proximity. We then examined how these nonverbal cues were associated with consumer purchase behavior and through what mechanisms in live-streaming settings. The results showed that each nonverbal cue was positively and significantly associated with consumer purchase behavior. Mediation tests further indicated that immersion significantly helped explain the relationships between nonverbal cues and consumer purchase behavior. From a process perspective, this study extends the range of constructs examined in live-streaming commerce and clarifies how nonverbal communication is associated with outcomes, offering practical implications for streamer training, camera setup, and content design.

Keywords: live-streaming commerce; nonverbal cues; immersion; purchase intention; multimodal analytics

Academic Editors: José Luís Mendes Loureiro Abrantes, Natália de Lima Figueiredo, Bruno Morgado Ferreira and Luís F. Martinez

Received: 11 February 2026
Revised: 26 March 2026
Accepted: 26 March 2026
Published: 29 March 2026

1. Introduction

Live-streaming commerce is an emerging sales model that uses online video live streams to integrate product presentation, interactive explanation, and online transactions [1]. As the live-streaming business model has matured and consumers purchasing habits have become more stable, live-streaming sales have shifted from a traffic channel characterized by episodic growth to a core vehicle for merchants routine operations [2]. In China, for example, sales generated through merchant-run live sessions exceeded RMB 430 billion in 2025, and merchant-run sessions accounted for 69% of all live-streaming commerce merchants [3]. Identifying how merchants can continuously attract consumers and stimulate consumption through manageable elements has become a pressing issue in live-streaming commerce research and practice.

Compared with traditional online selling formats that rely primarily on text-and-image information and customer service chats, live-streaming commerce enables real-time interaction. This interactivity provides consumers with more contextualized information during viewing and helps them form a more intuitive understanding of product performance [4]. Streamers often reduce consumers' perceived uncertainty through live demonstrations, firsthand trial, and ongoing explanation, which in turn facilitates the formation of product evaluations [5]. At the same time, live-streaming commerce also relies on a process-based experience to sustain consumers' ongoing attention and emotional engagement [6]. A large body of research indicates that consumers are more likely to enter a state of sustained focus, emotional involvement, and a diminished sense of time while watching live streams. This immersion experience significantly increases viewing value, lowers the likelihood of exit, and ultimately promotes consumer purchase behavior [7]. Therefore, understanding how immersion is activated and how it connects streamer presentation to consumer purchasing is a key pathway for explaining conversion efficiency in live-streaming commerce.

In live-streaming commerce, immersion is shaped not only by the transmission of verbal content but also, to a large extent, by streamers' nonverbal cues [8]. Nonverbal cues generally refer to communication channels that do not rely on semantic content, including facial expressions, gaze, gestures, body posture, spatial distance, and paralinguistic features [9]. Nonverbal behavior can convey interaction involvement and emotional states to audiences within a very short time, thereby influencing how information is processed and the intensity of the experience [10]. Recent studies have also shown significant associations between streamers' gesture use frequency, the extent of facial feature changes, and live-streaming consumer purchase behavior [11,12]. However, existing evidence remains limited in two ways. First, live-streaming commerce research overall still places greater emphasis on verbal strategies and script design [13], and discussions of nonverbal expression are relatively fragmented. Second, even when studies focus on nonverbal cues, they mostly examine more salient cues such as gaze, gestures, and facial expressions [14], while paying insufficient attention to head movements and bodily dynamics that are more processual and rhythmic. Because immersion is fundamentally about sustained attention and experiential involvement, it is more likely to be driven by high-frequency, continuous, and visually salient dynamic behavioral cues [15]. Without systematically identifying and testing these cues, our ability to explain the mechanisms of immersion in live-streaming commerce and its downstream conversion consequences will remain constrained.

Therefore, this study further places its focus on nonverbal behavioral cues that are more dynamic and processual in live-streaming settings. Specifically, this research focuses on four categories of nonverbal cues in live-streaming settings that are important yet have received limited scholarly attention: nodding frequency, forward lean, postural movement intensity, and camera proximity. It then examines their relationship with consumer purchase behavior. First, nodding is a common feedback behavior that can express understanding, agreement, and responsiveness without adding new information [16]. In live-streaming settings, viewers may interpret nodding as a cue of active response and interactional coordination, which may help maintain smooth interaction and sustain attention to the live-streaming content [17]. Second, forward lean is generally regarded as an approach-oriented posture that may signal greater attentional involvement and willingness to interact [18]. In live-stream viewing, this cue may shape viewers' perceptions of psychological distance and inform their judgments about the streamer's interaction tendency and communicative engagement [19]. Third, postural movement intensity reflects the streamer's behavioral energy and rhythmic variation during expression [20]. Such dynamic cues are more likely to attract attention and may convey greater expressive vitality and interactional tension [21]. Finally, camera proximity is an important visual presentation cue in live-streaming settings. By shaping visual distance and on-screen prominence, it may influence viewers' perceptions of psychological distance and attentional focus, thereby providing external conditions for the formation of immersion [22]. Overall, these cues capture nonverbal presentation characteristics in live-streaming settings that are readily perceptible, dynamically continuous, and relatively manageable, thus offering a perspective that is more closely aligned with the real live-streaming process for understanding how immersion forms.

To explain how the above nonverbal cues may shape viewers' attention allocation and social inference, and in turn relate to immersion and consumer purchase behavior, this study draws on Social Cognitive Theory as its core explanatory framework. Social Cognitive Theory suggests that individuals allocate attention to perceptible external behavioral cues and cognitively interpret them, thereby forming inferences about others' intentions, levels of interactional involvement, and behavioral motives. These inferences may further shape attitude formation and behavioral decision making [23]. In live-streaming commerce, consumers lack offline verification and sufficient two-way interaction. As a result, they rely more heavily on the behavioral cues that streamers display on camera to form judgments [24]. From the perspective of Social Cognitive Theory, dynamic nonverbal cues such as streamers' nodding, forward lean, postural movement, and camera distance serve as important sources of information that viewers use to allocate attention, understand the streamer's state, and make social inferences [5]. During sustained viewing, consumers may use these cues to form judgments about the streamer's interactional involvement, recommendation sincerity, and expressive state. These judgments may further shape their level of experiential involvement and, in turn, be associated with consumer purchase behavior. Accordingly, immersion may represent a key psychological mechanism linking streamers' nonverbal cues to consumer purchase behavior.

To test these inferences, this research examined merchant-run live sessions on Douyin and constructed a multimodal dataset based on 546 live sessions and 4600 product presentation segments matched to transaction outcomes. Using an automated computer-vision–based framework, we extracted segment-level nonverbal cues and linked them with minute-level operational data. This design allowed us to test how streamers' dynamic nonverbal cues relate to immersion and consumer purchase behavior within the same unit of analysis.

2. Literature Review

2.1. Streamer Behavior and Consumer Decisions

As the central agent who conveys product information and interacts with viewers in live-streaming commerce, the streamer is widely viewed as an important driver of consumers purchase decisions. Prior research can be grouped into three broad categories. The first stream examines streamers' appearance and attractiveness. Evidence suggests that streamers' attractiveness, defined as an overall evaluation of the streamers' looks, personality, and talent, increases viewing enjoyment and strengthens purchase intention [25]. In addition, streamers' facial attractiveness can influence how viewers allocate attention and, to some extent, reduce psychological resistance to the shopping context, thereby increasing consumers purchase intention [26]. The second stream focuses on the characteristics of streamers interactive content. Yang et al. (2023) found that when streamers use a stronger social-interaction-oriented communication style during live streaming, it significantly affects viewers purchasing and tipping behavior [27]. Using Douyin live-streaming data, Yang and Wang (2025) showed that streamers who maintain high-frequency interaction with viewers and engage in more diverse communication behaviors tend to be more popular during live streams [28]. Gu et al. (2023) found that information-rich live-streaming content is often accompanied by stronger enjoyment and perceived social support [29]. As a result, consumers are more likely to have a memorable viewing experience and to increase their intention to continue participating. Research also indicates that the level of professional competence displayed by streamers during interaction, as well as the quality of the information content, meaningfully shapes consumers attitude formation and purchase decisions [30]. Further work suggests that interaction quality in live streaming is not a single construct. It can be decomposed into responsiveness, professionalism, informativeness, and personalization, and these dimensions show significant associations with consumers emotional experiences and purchase intention [31]. The third stream emphasizes fit-related characteristics, including the congruence between the streamer and the product and the match between the streamer and consumers. Prior research indicates that when streamer—product fit is high, viewers are more likely to develop favorable perceptions of the streamer's attractiveness and credibility [32]. Live-streaming interaction not only transmits information but also serves as a process through which social identification is gradually constructed. When viewers perceive that the values communicated by the streamer align with their own, they are more likely to identify with the streamer and the explanations provided, which in turn influences attitude formation and behavioral responses [33].

In summary, existing research has primarily examined streamer behavior in live-streaming commerce from the perspective of interactive content, with a focus on verbal cues such as language use and interaction frequency and their effects on consumer attitudes and purchasing. Compared with the extensive attention given to verbal cues, research has paid limited attention to the nonverbal cues that streamers display during live streaming.

2.2. Nonverbal Cues

In communication and media research, nonverbal cues are widely regarded as an indispensable component of human interaction. They are expressed through multiple channels, including facial expressions, bodily movements, body posture, and vocal paralinguistic features [34–36]. Nonverbal cues often convey emotions and psychological states in an unconscious manner and serve multiple functions in communication. Prior research shows that nonverbal cues shape behavioral decisions by influencing social perception and cognitive inference processes, which in turn affect judgments about others attitudes, intentions, and credibility [10,37].

From a functional perspective, different types of nonverbal cues play distinct roles in interpersonal exchange. Research finds that gestures not only reinforce the expression of key information but also effectively direct audience attention, thereby improving the efficiency of information processing during communication. In highly interactive media contexts such as live streaming, these benefits are particularly pronounced when gestures align with verbal content [38]. Eye contact is another critical nonverbal cue. It is typically viewed as a bidirectional interaction signal. Both actively looking at others and being looked at tend to be accompanied by physiological arousal, which gradually manifests in individuals emotional experiences and behavioral responses during interaction [39]. With respect to facial nonverbal cues, Todorov et al. (2015) systematically showed that people quickly infer others credibility and intentions from cues such as facial expressions [40]. Related work further suggests that smiling, as a specific facial expression, can foster trust and thereby increase a communicator's perceived warmth and approachability [41]. In addition, body posture is considered an important nonverbal cue. It may strengthen viewers mental imagery and product evaluations and promote consumer purchase behavior in combination with verbal information [12].

https://doi.org/10.3390/jtaer21040106

In terms of measurement, advances in computational methods have made the quantification of nonverbal cues increasingly feasible and have enabled researchers to capture nonverbal signals more efficiently in video media such as advertising and live streaming. Prior research has used automated facial coding, expression recognition, and facial tracking techniques to continuously measure viewers' emotional responses and to assess how different video content is associated with differences in viewing intention and market performance [42,43]. For bodily cues, Kim et al. (2021) used OpenPose to extract full-body skeletal keypoints and computed joint angles of the trunk and limbs to quantify posture during task performance [44]. For head movement recognition, Kujani and Kumar (2023) used a convolutional neural network-based transfer learning approach to automatically identify behaviors such as nodding and head turning in real-time video, thereby converting head movement patterns into quantifiable nonverbal indicators [45].

2.3. Social Cognitive Theory

Social cognitive theory (SCT) emphasizes that individuals are not passive recipients of external stimuli. Instead, in social contexts, they continuously attend to and process others observable behaviors to form understanding and judgments about interaction partners, and they develop relatively stable attitudinal orientations and behavioral tendencies on this basis [46]. External cues do not translate directly into behavioral outcomes. Rather, their effects are translated into individuals decisions and actions through internal cognitive processes such as attention allocation, meaning interpretation, and social inference [47]. More broadly, SCT is grounded in triadic reciprocal determinism, which posits mutual interaction and reinforcing feedback loops among environmental inputs, individual psychological processes, and behavioral responses. This framework explains how others behaviors in communication settings shape audiences interaction and choice [48].

Live-streaming commerce provides a highly prototypical setting for applying SCT. A live-streaming room is a socially interactive space mediated by the camera. Because consumers have limited opportunities for offline verification, they are more likely to construct cognitions and make value judgments by continuously observing the streamer and the interaction context [49]. In this process, viewers not only process product information itself but also continuously integrate streamers' behavior, the interaction atmosphere, and emotional cues [50]. Based on these inputs, they form inferences about the streamer's credibility, recommendation intent, and level of interactional involvement, which may ultimately be reflected in differences in viewing preferences, participation behavior, and purchase decisions [51]. Because live-stream viewing is a process of sustained attention and experiential involvement that unfolds through ongoing observation [52], the role of streamers' behavioral cues may extend beyond viewers' judgments of the streamer's interactional involvement and recommendation intent. These cues may also shape viewers' sustained attention to the live-streaming content, the coherence of their understanding, and their level of situational involvement [53]. Such sustained attentional engagement and experiential absorption can be regarded as important psychological foundations for the formation of immersion [54]. Joo and Yang (2025) showed that in live-streaming commerce, stronger immersion helps intensify consumers information processing during viewing and interaction and further increases purchase intention by strengthening cognitive evaluations and eliciting positive emotional responses [55].

Within the SCT framework, the nonverbal cues that streamers display on camera constitute a key source of information for audiences observational learning and social inference [11]. Compared with verbal content, nonverbal cues are typically more immediate and more tightly tied to the situation. They can shape attention allocation, credibility judgments, and outcome expectations under lower cognitive load, thereby influencing experiential involvement and subsequent behavioral responses. Following this theoretical logic, research has begun to empirically test specific nonverbal cues. For example, Liu and Zhao (2025) found that a higher level of nonverbal coordination between two streamers is associated with stronger user responses and consumer purchase behavior in live-streaming settings [12]. Zhang et al. (2025) noted that viewers form social judgments based on streamers' facial features, which may further shape their evaluations of both the streamer and the product [56]. Xian et al. (2025) incorporated influencers facial attractiveness and body-movement characteristics into their model and find that nonverbal cues significantly moderate the persuasive effects of verbal appeals and translate into consumer purchase behavior [57].

In summary, SCT provides a clear theoretical pathway for explaining how nonverbal cues in live-streaming commerce shape consumer responses. Viewers form social inferences by observing the streamer's visible behaviors and how they are presented. These inferences alter attention investment and experiential involvement, and ultimately affect consumer purchase behavior. Although prior research has documented the importance of some nonverbal cues, it has not systematically tested dynamic behavioral cues that are more processual and more manageable. Building on this gap, this study further focuses on nodding frequency, forward lean, postural movement intensity, and camera proximity. Within the SCT framework, we aim to show how these cues influence consumer purchase behavior through immersion. In doing so, we expand the set of variables examined in research on nonverbal communication in live-streaming commerce and deepen understanding of the psychological mechanisms underlying conversion in live streaming.

It should be noted that the salience and communicative meaning of nonverbal cues may vary across cultural backgrounds and interaction contexts. Prior research has shown that audiences from different cultures rely on different nonverbal cues when identifying and interpreting others' responses, and that social context also shapes the use and intensity of nonverbal behavior [58]. At the same time, live-streaming commerce is characterized by real-time communication and a strong emphasis on the streamer's on-screen performance, such that consumers continuously receive real-time information and observe the streamer's performance throughout the decision-making process [59]. Accordingly, the mechanism through which nonverbal cues operate in this study is discussed primarily in the context of China's live-streaming commerce market.

3. Hypotheses and Research Model

3.1. Nodding

Nodding is a prototypical nonverbal cue that is commonly used to express agreement, cooperation, and alignment with the rhythm of an interaction. It is a highly prevalent feedback signal in interpersonal communication and can be quickly recognized [60]. Prior research suggests that nodding is often interpreted as an overt marker of approach motivation and a positive interaction orientation, which enhances perceptions of an interaction partner's warmth and positivity [61]. In live-streaming commerce, a streamer's nodding while explaining product benefits, responding to chat questions, or acknowledging viewers' emotions conveys an immediate sense of responsiveness. Without adding informational content, it provides viewers with a sense of being noticed and responded to, thereby strengthening the continuity of interaction and stabilizing the rhythm of the viewing experience. Viewers may further interpret nodding as a sign that the streamer is actively responding and maintaining the rhythm of the interaction. Such judgments may promote a smoother interaction process and help sustain viewers' attention to the live-streaming content [17]. Drawing on social cognitive theory, viewers treat nodding as a salient visible behavioral cue and use it to form social inferences about the streamer's interaction involvement and attitudinal orientation [62]. When nodding is more frequent, viewers are more likely to infer that the streamer is actively engaged in the interaction and is willing to stay in sync with the audience. This perception may help maintain viewers' attention to the live-streaming content and make them more likely to enter an immersion state characterized by sustained attention and experiential involvement. Stronger immersion means that viewers maintain higher levels of attentional investment and emotional involvement during viewing [54]. They are less likely to discontinue viewing and more likely to complete the behavioral transition from interest to decision and then to order placement within the focal context [55]. Accordingly, nodding may be related to consumer purchase behavior and may also be associated with consumer purchase behavior through immersion. We therefore propose:

H1. *Streamers' nodding frequency has a positive effect on consumer purchase behavior.*

H2. *Immersion mediates the relationship between streamers' nodding frequency and consumer purchase behavior.*

3.2. Postural Movement

Postural movement intensity reflects the magnitude of changes in a streamer's body posture and movements during live-streaming commerce [57]. In live-streaming contexts, higher postural movement intensity is typically accompanied by stronger expressive tension and more pronounced rhythmic shifts. Such dynamic behavioral cues are more likely to receive prioritized processing in viewers' attention systems [63], thereby enhancing sustained attention and emotional arousal during viewing [64]. Related research also suggests that the intensity and contagion of physical expression strengthen audiences' emotional responses and engagement and further shape consumption responses [21], thereby increasing purchase intention. When viewers observe high postural movement intensity from a streamer, they are more likely to interpret it as evidence of stronger commitment to the recommendation, more forceful expression, and a willingness to maintain interaction over time. These perceptions may help maintain viewers' engagement with the live-streaming content and make them more likely to enter an immersion state. Under immersion, viewers are more likely to stay longer, interact more, and exhibit lower exit rates. These behavioral trajectories are typically consistent with a higher likelihood of order placement [52]. We therefore proposed:

H3. *Streamers' postural movement intensity has a positive effect on consumer purchase behavior.*

H4. *Immersion mediates the relationship between streamers' postural movement intensity and consumer purchase behavior.*

3.3. Camera Proximity

Camera proximity refers to the visual distance and the degree of camera closeness presented in a live-streaming view [65]. It is a highly manageable visual presentation cue in live-streaming settings and one that may have a direct bearing on viewers' experience. A closer camera view typically creates a more concrete and more psychologically proximal experience, making it easier for audiences to form vivid representations and immediate judgments, thereby strengthening behavioral tendencies [66]. Within the SCT framework, camera distance is an important situational cue that viewers use to infer interaction accessibility and psychological distance [67]. When the camera is closer, viewers are more likely to experience a sense of being brought closer. They are more likely to focus their attention on the streamer's delivery and product details. The viewing process becomes less disrupted and more continuous, making it easier to enter an immersion state. Stronger immersion implies greater situational involvement and a lower likelihood of interruption, which makes it more likely that purchase decisions are completed within the live-streaming context [68]. We therefore proposed:

H5. *Streamers' camera proximity has a positive effect on consumer purchase behavior.*

H6. *Immersion mediates the relationship between streamers' camera proximity and consume purchase behavior.*

3.4. Gestures

Gestures are an important form of nonverbal expression in live-streaming communication. They can highlight key information, guide viewers' attention, and increase clarity and persuasive effectiveness [38]. Prior research indicates that gestures not only heighten audiences' attention but that gesture intensity is also closely related to expressive appeal and the effectiveness of information processing [69]. In live-streaming commerce, the pace of information delivery is fast and the content is dense. Viewers process a large amount of product benefits, promotion information, and interactive content within a short time, which makes attention fluctuations and rising information load more likely [70]. In such situations, more salient and dynamic gesture cues can help viewers capture the key points and reduce comprehension costs, while also stabilizing attention during viewing [71]. From a social cognitive theory perspective, viewers are likely to interpret stronger gestures as an overt signal that the streamer is emphasizing core information and expressing a stronger recommendation intent, thereby increasing their tendency to adopt the recommended content [72]. At the same time, the attention guidance and rhythmic reinforcement generated by stronger gestures may further enhance viewers' sustained attention to the live-streaming content and their sense of participation, which may in turn be reflected in a higher level of immersion. A higher level of immersion may, in turn, make viewers more likely to remain engaged and interact, and may also be associated with consumer purchase behavior in the live-streaming context. We therefore proposed:

H7. *Streamers' gesture intensity has a positive effect on consumer purchase behavior.*

H8. *Immersion mediates the relationship between streamers' gesture intensity and consumer purchase behavior.*

3.5. Forward Lean

Forward lean is a postural cue with clear relational meaning. It typically communicates stronger engagement, approach, and willingness to interact [73]. Research on relational meaning in nonverbal behavior suggests that body orientation and the degree of leaning forward significantly shape relational judgments and trust tendencies between interaction partners [37]. In live-streaming contexts, forward lean is not only a postural expression. It also changes the on-screen body proportion and the location of attentional focus, making the streamer's delivery feel more psychologically close and thereby increasing viewers' perceived interaction accessibility [74]. Drawing on social cognitive theory, viewers treat forward lean as a visible cue for judging the streamer's interaction involvement and recommendation sincerity. When the streamer leans forward more, viewers are more likely to infer that the streamer values communicative feedback and is willing to build a connection with the audience. More importantly, the visual focus and psychological closeness induced by forward lean can reduce psychological distance and the likelihood of distraction during viewing. This may help viewers sustain attention and experiential involvement more easily, thereby strengthening immersion. Stronger immersion may help viewers maintain continued viewing and interaction participation, and may further be associated with their consumer purchase behavior [75]. We therefore proposed:

H9. *Streamers' forward lean has a positive effect on consumer purchase behavior.*

H10. *Immersion mediates the relationship between streamers' forward lean and consumer purchase behavior.*

Based on these hypotheses, we proposed a conceptual model, as shown in Figure 1.

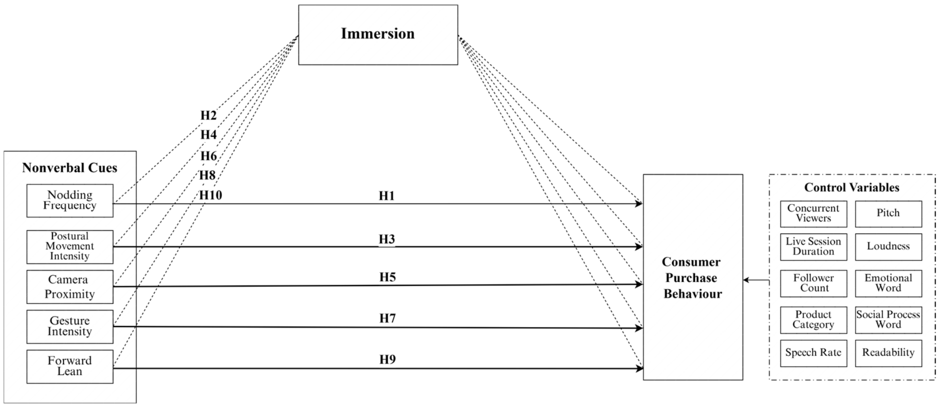

Figure 1. Conceptual Model of This Study.

4. Method

4.1. Data Processing

This research used commercial live-streaming sessions on Douyin as the empirical setting. As one of the largest and most active short-video and live-streaming commerce platforms in China, the Douyin app reached 948 million monthly active users. It aggregates a large and stable supply of live-streaming transactions as well as high-frequency viewing and interaction behaviors, making its live-streaming data highly representative [76]. To ensure data availability and consistent measurement, we obtained structured live-room metrics from a third-party Douyin data service platform and simultaneously archived the live-streaming video content. This process enabled us to construct a multimodal dataset that could be aligned within the same unit of analysis.

For data collection, we wrote a real-time monitoring and scraping program in Python (3.10) and continuously tracked and collected Douyin commercial live-streaming data during March 2025. To reduce systematic effects of time-of-day differences in audience supply on sales, data collection was restricted to the high-activity evening window from 19:00 to 22:00. Prior research has generally suggested that user density is higher during this period and that viewing and interaction are more concentrated, making it more reflective of typical live-streaming operations [77]. During the study period, we collected 546 commercial live-streaming sessions, covering 120 streamers.

For data processing, we constructed the analytic dataset at the product-presentation segment level and aligned multi-source data at the same granularity. Because a single live session typically contains sequential demonstrations of multiple products, using the entire session as the unit of analysis would conflate variation across product stages with their corresponding transaction outcomes and weaken identification. We therefore segmented each live-session replay based on product-switch boundaries, marking the start of a segment when a focal product became the primary item being demonstrated and the end when the streamer switched to the next product. For each segment, we recorded start and end timestamps and used them to align video-based measures with platform-provided segment-level outcomes.

To ensure segment quality and matching reliability, we applied a set of cleaning and filtering rules. We retained only segments that could be uniquely matched to a platform product identifier and a corresponding segment-level sales record, contained complete video and audio streams, and met a minimum duration requirement of 10 s to provide sufficient behavioral signal for automated extraction. We excluded segments with missing values in key fields (e.g., sales, timestamps, average watch time, concurrent viewers, or product identifiers), inconsistent temporal information (start time later than end time), or zero/negative duration. To reduce measurement noise in the computer-vision pipeline, we further required a minimum valid-frame coverage rate of 90%, defined as the proportion of frames with successful keypoint detection; segments below this threshold were removed because they typically involved severe occlusion, low visual quality, or extended off-camera periods. After these procedures, the final analytic sample consisted of 4600 valid product-presentation segments with one-to-one matching to product-level sales outcomes.

In addition to sales, we also obtained minute-level structured operational data from the live-streaming process and aligned it with the product presentation segments. Specifically, within each live session, platform-provided structured information was used to extract key live-session and product metrics, including live-stream duration, number of concurrent viewers, average watch time, and product-level sales. We also collected streamers baseline attributes and operational characteristics. We then aggregated or averaged these minute-level indicators within each segment window to construct process features for the corresponding product presentation segment. This approach allowed us to observe streamers' behavioral cues and the real-time operational environment at the same product level. Ultimately, we constructed a multimodal dataset with the product presentation segment as the core unit of observation. It included minute-level structured process data and product sales outcomes, as well as unstructured information such as video, audio, and transcribed text. This dataset provided the foundation for subsequent multimodal analyses of visual, audio, and textual data. The data processing workflow is shown in Figure 2.

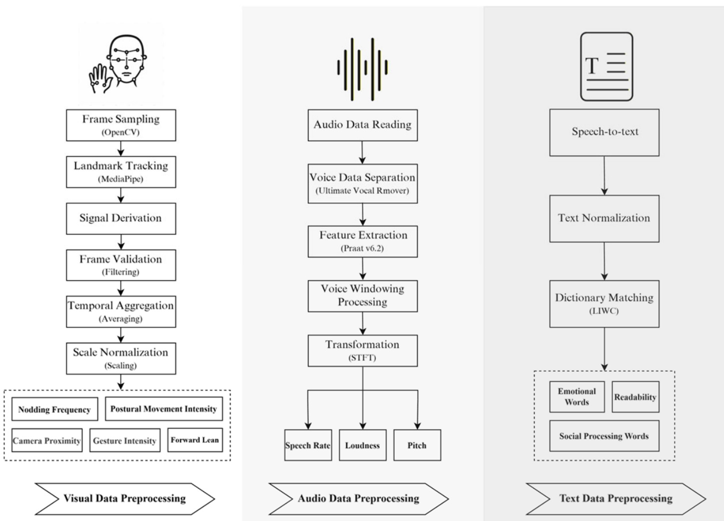

Figure 2. Data processing flowchart.

4.2. Variable Measurement

To quantify nonverbal cues, we used a computer-vision-based automated recognition framework that converted facial and hand keypoint information in live-streaming videos into structured measures for econometric analysis. Specifically, we first used OpenCV to load the videos and extracted frames at a fixed sampling rate, transforming the continuous video stream into a sequence of frames. We then invoked MediaPipe FaceMesh and Hands to detect the streamer's facial and hand keypoints and tracked them continuously over time. Based on these outputs, we constructed frame-level behavioral signals using the spatial geometric relationships among keypoints and their displacement changes across frames. Finally, we aggregated the frame-level signals within each product presentation segment window and standardized them by incorporating segment duration and on-screen scale, thereby producing segment-level variables. To reduce measurement noise caused by occlusion, momentary jitter, or detection failures, we computed the measures using only valid detected frames. We also aggregated key indicators over time at the segment level using means or cumulative values, so that the variables captured stable behavioral characteristics rather than incidental fluctuations.

4.2.1. Independent Variable

The five nonverbal cues examined in this study, namely nodding frequency, gesture intensity, postural movement intensity, camera proximity, and forward lean, represent five distinct yet potentially related dynamic behavioral cues. Based on the aspects of nonverbal presentation they reflect and the logic underlying their measurement, these cues can be organized into three broader categories: expressive cues (nodding frequency, gesture intensity, and postural movement intensity), proximity-related cues (camera proximity), and attentional or interactional signaling cues (forward lean) [10]. This categorization facilitates a clearer understanding of the five nonverbal cues.

1. Nodding frequency

Nodding frequency captured the intensity of feedback-related head movements that the streamer displayed during explanation and interaction. Based on FaceMesh outputs, we constructed a time-series signal of head pitch changes using keypoints such as the nose tip and both eyes, and we applied scale normalization to reduce distortions caused by camera distance and on-screen scaling [78]. We then identified rhythmic patterns in the pitch signal that exhibited direction reversals within a short time window and were accompanied by noticeable velocity changes. These patterns were used to detect nodding events and count their occurrences. Finally, we divided the total number of nods by segment duration to obtain a per-unit-time nodding frequency measure, ensuring comparability across product presentation segments.

2. Gesture intensity

Gesture intensity reflected the amplitude of the streamer's movements and the extent to which information was emphasized during delivery. We used MediaPipe Hands to detect and track sequences of hand keypoints. We then operationalized gesture motion as the cumulative displacement of hand keypoints between adjacent frames [79]. To ensure that the measure was not systematically affected by differences in resolution or shooting distance, we applied on-screen scale normalization to the displacement values [80]. At the segment level, we standardized the cumulative hand displacement by segment duration to obtain gesture intensity per unit time. This approach more accurately captured the relative activeness of gesture expression across product presentation segments.

3. Postural movement intensity

Postural movement intensity measured the streamer's overall movement activeness and expressive energy in front of the camera. Because full-body keypoints in live-streaming settings can be unstable due to framing and occlusion, we used stable facial keypoints as a proxy for upper-body movement. We captured macro-level movement amplitude by tracking the displacement changes in these points between adjacent frames and cumulatively summing the displacement within each segment window to obtain total postural movement [81]. We then standardized this total by segment duration to derive postural movement intensity per unit time, ensuring consistent measurement meaning across segments with different presentation lengths [80].

4. Camera proximity

Camera proximity captured the visual distance between the streamer and the camera as presented in the live-streaming view. This indicator emphasizes perceived closeness at the level of screen composition and visual presentation, primarily reflecting the camera-based presentation effect itself rather than the streamer's actively adopted approach-oriented posture. Based on the set of facial keypoints from FaceMesh, we computed the geometric scale of the face in the frame and used the face's relative on-screen coverage as a proxy for proximity [82]. After computing this measure at the frame level, we averaged it across valid detected frames within each product presentation segment to obtain segment-level camera proximity. This aggregation smoothed short-term fluctuations and yielded a more stable representation of the camera presentation level.

5. Forward lean

Forward lean captured the intensity of the streamer's approach-oriented posture during live streaming. This indicator is intended to reflect the streamer's active tendency to lean forward during explanation and interaction, thereby capturing interactional involvement and approach-oriented intent more directly. Forward lean is typically reflected in both changes in apparent facial scale and shifts in the position of the face within the frame [83]. Therefore, this study does not identify forward lean solely on the basis of facial scale variation. Instead, it combines information on facial keypoint coverage and the vertical distribution of keypoints to distinguish between two different situations, a closer camera and the streamer's active forward movement, and then integrates them to construct the forward lean score. At the segment level, we averaged the forward-lean scores across valid detected frames to obtain the forward lean measure, which reflected the streamer's overall leaning tendency and interaction posture intensity during the focal product presentation stage [84].

To assess the reliability of the computer-vision-based measures, we conducted a manual validation on a random subsample of 200 product-presentation segments. Two trained coders independently annotated key behaviors following a protocol aligned with our computational definitions (e.g., nodding events, forward-lean posture, and camera proximity categories). Inter-coder reliability for segment-level summaries was satisfactory (ICC = 0.82). We then compared manual annotations with the automated outputs. For nodding frequency, the correlation between the algorithmic count (per minute) and the manual count was $r = 0.758$ ($p < 0.001$), and event-level agreement yielded precision = 0.88, recall = 0.89, and F1 = 0.89. For forward lean and camera proximity, agreement between manual labels and automated scores was high (accuracy = 0.91). These checks suggest that the automated pipeline captures the targeted nonverbal behaviors with acceptable accuracy for large-scale analysis.

To address potential overlap among the five nonverbal indicators, we assessed their bivariate correlations and multicollinearity diagnostics. The pairwise correlations were generally low to moderate, suggesting that the indicators capture related but distinct aspects of nonverbal presentation ($r < 0.3$). As expected, forward lean and camera proximity were positively correlated because both partially reflect changes in on-screen facial scale ($r = 0.225$); however, the correlation remained well below levels typically associated with problematic redundancy, and the remaining indicator pairs showed weaker associations. Consistent with this pattern, multicollinearity was not a concern in the full specification: the variance inflation factors were uniformly small (VIFs close to 1, with the maximum VIF = 1.28), indicating that the five indicators provide separable explanatory content when entered jointly.

4.2.2. Dependent Variable

The dependent variable in this study was the live-streaming sales associated with each product presentation segment. We use segment-level sales as an objective behavioral indicator of consumer purchase behavior because, in live-streaming commerce, purchases are executed and recorded directly within the live room during or immediately following a focal product presentation. As a result, sales capture realized conversion rather than self-reported intention and have been widely used in prior live-streaming and digital commerce research as a valid proxy for purchasing outcomes. After completing video segmentation and quality screening, we retained 4600 valid product presentation segment observations. At the product level, we matched each segment one-to-one with the corresponding platform sales record to construct the segment-level sales measure. This procedure ensured that the purchase outcome and the corresponding behavioral cues were aligned within the same unit of analysis, thereby improving the accuracy of subsequent econometric identification.

4.2.3. Mediating Variable

In this study, average watch time is used as a behavioral proxy related to immersion, but this indicator is not equivalent to immersion as a psychological experience in itself. At the behavioral level, immersion is often reflected in stronger sustained attention and a lower tendency to exit [85]. Accordingly, watch time may, to some extent, reflect viewers' level of experiential involvement during the product presentation stage [86].

It should be noted, however, that although watch time can capture immersion-related engagement to a certain extent, it may also be influenced by platform algorithms, overall live-stream duration, and viewers' continued presence while waiting for discount information or product logistics details. In this study, the minute-level watch time data provided by the platform are aligned with each segment's time window and then aggregated or averaged within the segment to obtain the average watch time for that segment, which is used to characterize the level of immersion.

4.2.4. Control Variables

To reduce confounding influences from the live-streaming operational environment, streamer heterogeneity, and product attributes on sales, we included multiple sets of control variables in the model. First, we controlled for concurrent viewers, live session duration, follower count, and product category. Concurrent viewers captured real-time room popularity and the potential base of conversions, reflecting differences in transaction opportunities driven by changes in audience size [87]. Live session duration reflected the supply of content and the intensity of operational investment [88]. Follower count controlled for stable streamer-level differences [89]. Product category controlled for systematic differences across products in decision complexity and demand structure [90], thereby improving the robustness of the core estimates.

In addition, we further controlled for the audio and text features of each live-streaming segment to reduce potential interference from differences in verbal expression on the relationship between nonverbal cues and sales. Persuasion in live-streaming commerce relies heavily on vocal delivery and scripted talk. Speech rate and acoustic properties can directly affect attention maintenance and emotional arousal [91], and emotional word use and social process word use in text may co-vary with streamers nonverbal presentation [92].

Accordingly, controlling for both audio and text features helped more robustly identify the net effects of nonverbal cues.

For audio measurement, we first extracted the audio track from the video corresponding to each product presentation segment and used a source-separation approach to split the vocal signal from background music. This procedure reduced interference from accompaniment, ambient noise, and mixing effects in acoustic feature extraction. Specifically, we used Ultimate Vocal Remover (UVR GUI, v5.6) for two-source separation. This tool is based on deep-learning source-separation models that perform spectral decomposition and mask estimation in the time-frequency domain, producing a separated vocal track and a nonvocal track [93]. We used the separated vocal track for subsequent acoustic analyses and retained the nonvocal track to support extended measures such as background-music rhythm. Next, we input the vocal signal into Praat (v6.2) to extract acoustic features [94]. Following standard speech-processing procedures, Praat segmented continuous speech into short-time frames of 20 to 30 milliseconds with 50% overlap. It then applied windowing and a short-time Fourier transform to obtain stable time—frequency features, which were used to extract the core acoustic measures of speech rate, pitch, and loudness. Finally, we averaged these acoustic measures within each segment window to construct segment-level audio control variables.

For text-feature measurement, we transcribed the vocal audio in each segment into text and performed standardized cleaning to ensure comparability across segments. We then used LIWC (2022) to conduct dictionary matching and proportion calculations. This allowed us to extract the shares of emotional word use and social process word use in the script, capturing the intensity of emotional expression and the orientation toward social interaction [95]. In addition, text readability was measured by the share of easy-to-understand words. A higher proportion of such words indicates lower linguistic complexity and lower information-processing costs [96]. All text measures were computed at the segment level and entered the regression models as control variables. This approach allowed us to control for the effects of both vocal delivery and language content differences on sales in the econometric tests.

4.3. Model Specification

In our dataset, consumer behavior indicators such as sales and average watch time typically exhibited strong right skewness and long-tail distributions. Using raw values in linear estimation would be sensitive to extreme observations and could induce heteroskedastic and non-normal errors, thereby weakening the robustness of statistical inference. To mitigate distributional skewness and improve comparability across observations, we applied a log transformation $\ln(1 + x)$ to the variables, which preserved zero-valued observations while compressing extreme values. This treatment is consistent with common empirical practices for handling long-tail behavioral data in marketing and information systems research [97].

For main-effect tests, we estimated the effects of nonverbal cues on sales using linear regression models with fixed effects. All models controlled for operational variables and included live session fixed effects and product category fixed effects to absorb unobserved systematic differences at the live-session and category levels. For mediation tests, we used the product-of-coefficients approach within a two-step regression framework [98]. We first estimated the effect of nonverbal cues on immersion to obtain the a-path coefficient. We then added immersion to the sales equation to obtain the b-path coefficient and computed the indirect effect accordingly. Confidence intervals for the indirect effect were constructed using the delta method based on large-sample approximation. Tests for multicollinearity showed that VIF values were low across models (VIF < 5; ref. [99]), suggesting no substantial risk of collinearity in coefficient estimation.

5. Results

5.1. Main Effects

Table 1 reports the regression results for seven models. Model 1 was the baseline model. Models 2 through 6 added the five types of nonverbal cues separately, and Model 7 included all nonverbal cues simultaneously to identify their net effects. The results showed that the five nonverbal cues examined in this study explained significant variation in sales, and the conclusions remained robust after controlling for live session fixed effects and product category fixed effects. The baseline model had an R^2 of 0.24. When a single nonverbal cue was added, the R^2 increased to between 0.28 and 0.32. When all five cues were included in Model 7, the R^2 further increased to 0.39, indicating that nonverbal cues provided substantial incremental explanatory power for sales beyond the control variables and fixed effects.

Table 1. Regression Analysis of the Effects of Nonverbal Cues on Sales.

Variable	Model 1 Baseline	Model 2 Nodding	Model 3 Gesture	Model 4 Movement	Model 5 Proximity	Model 6 Lean	Model 7 Full Model
ln(1 + Nodding)		0.16 *** (0.03)					0.17 *** (0.05)
ln(1 + Gesture)			0.22 * (0.09)				0.16 ** (0.06)
ln(1 + Movement)				0.21 *** (0.06)			0.18 ** (0.06)
ln(1 + Proximity)					0.16 ** (0.05)		0.14 *** (0.04)
ln(1 + Lean)						0.13 * (0.06)	0.12 ** (0.04)
Control variables	YES	YES	YES	YES	YES	YES	YES
Fixed effect (Live session)	YES	YES	YES	YES	YES	YES	YES
Fixed effect (Product category)	YES	YES	YES	YES	YES	YES	YES
VIF	1.15	1.15	1.26	1.26	1.26	1.26	1.28
N	4600	4600	4600	4600	4600	4600	4600
R^2	0.24	0.31	0.32	0.28	0.30	0.29	0.39

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors are in parentheses.

First, nodding frequency exhibited a stable positive relationship with sales. In Model 2 of Table 1, the coefficient for nodding frequency was $\beta = 0.16$ ($SE = 0.03$, $p < 0.001$), indicating that segments with more frequent nodding were associated with higher sales. In the full Model 7, the coefficient for nodding frequency was $\beta = 0.17$ ($SE = 0.05$, $p < 0.001$). The sign and significance remained consistent, suggesting that the effect of nodding frequency on sales was not explained away by other nonverbal cues. These results supported H1.

Postural movement intensity and camera proximity also showed significant positive effects, and both remained robust in the full model. In Model 4 of Table 1, the coefficient for postural movement intensity was $\beta = 0.21$ ($SE = 0.06$, $p < 0.001$). In the full Model 7, the coefficient was $\beta = 0.18$ ($SE = 0.06$, $p = 0.003$), remaining significantly positive and supporting H3. For camera proximity, the coefficient in Model 5 was $\beta = 0.16$ ($SE = 0.05$, $p = 0.001$). In the full Model 7, it was $\beta = 0.14$ ($SE = 0.04$, $p < 0.001$), indicating an independent explanatory contribution after simultaneously controlling for the other cues and supporting H5.

Gesture intensity and forward lean were also significantly and positively associated with sales. In Model 3 of Table 1, the coefficient for gesture intensity was $\beta = 0.22$ ($SE = 0.09$, $p = 0.015$). In the full Model 7, it was $\beta = 0.16$ ($SE = 0.06$, $p = 0.008$), indicating that the gesture effect remained significant after controlling for multiple cues and supporting H7. For forward lean, the coefficient in Model 6 was $\beta = 0.13$ ($SE = 0.06$, $p = 0.030$). In the full Model 7, it was $\beta = 0.12$ ($SE = 0.04$, $p = 0.003$), again showing a robust positive association and supporting H9.

5.2. Mediation Results

Next, we used the product-of-coefficients approach to test the mediating effect of immersion. Under the classic mediation logic, mediation analysis first requires that the independent variables significantly predict the mediator. We therefore first examined the effects of nonverbal cues on immersion. Immersion is operationalized using average watch time as a behavioral proxy. Because average watch time in live-streaming settings also exhibited right-skewed, long-tail distributions, we applied a log transformation ln (1+x) and used ln(1+WatchTime) as the dependent variable in the regression models. This treatment reduced the influence of extreme values and improved the robustness and comparability of the estimates.

In terms of model specification, we followed the main-effect estimation framework and constructed the same stepwise model sequence as in Table 1. All models included control variables, live session fixed effects, and product category fixed effects to absorb differences in the operational environment and other unobserved systematic factors at the live-session and category levels. It should be noted that nonverbal cues, average watch time, and sales outcomes are all measured at the same product presentation segment level. Therefore, the mediation analysis in this study is better understood as an examination of within-segment process relationships rather than a causal mediation interpretation under a strict temporal sequence. Model 1 was the baseline model and included only the control variables and fixed effects. Models 2 through 6 added nodding frequency, gesture intensity, postural movement intensity, camera proximity, and forward lean one at a time to test each nonverbal cue's unique explanatory power for viewing retention. Model 7 was the full model and included all five cues simultaneously to identify their net effects under mutual controls.

As shown in Table 2, the full Model 7 indicated that nodding frequency was positively associated with average watch time, with a coefficient of $\alpha = 0.21$ ($SE = 0.08$, $p = 0.009$). The corresponding coefficients were $\alpha = 0.25$ ($SE = 0.10$, $p = 0.012$) for gesture intensity, $\alpha = 0.16$ ($SE = 0.06$, $p = 0.008$) for postural movement intensity, $\alpha = 0.16$ ($SE = 0.07$, $p = 0.022$) for camera proximity, and $\alpha = 0.14$ ($SE = 0.05$, $p = 0.005$) for forward lean. Across models, R^2 increased from 0.18 in the baseline model to 0.31 in the full model, indicating that nonverbal cues significantly improved the explanatory power for average watch time. These results provided the prerequisite mechanism evidence for H2, H4, H6, H8, and H10.

Furthermore, after average watch time was included in the sales equation, it exhibited a significant positive effect on sales. In Model 10 of Table 3, the coefficient for immersion was $\rho = 0.32$ ($SE = 0.10$, $p = 0.001$), indicating that segments with longer average watch time were associated with higher sales.

Finally, Table 4 reports the indirect effects and confidence intervals. We computed the indirect effect using the product-of-coefficients approach, $IE = \alpha \times \rho$, and constructed 95% confidence intervals using the delta method. The results showed that the indirect effect for nodding frequency was 0.07, with a 95% CI of [0.0023, 0.1321], and the proportion mediated was 0.395. The indirect effect for gesture intensity was 0.08, with a 95% CI of [0.0004, 0.1596], and the proportion mediated was 0.500. The indirect effect for postural movement intensity was 0.05, with a 95% CI of [0.0022, 0.1002], and the proportion mediated was 0.284. The indirect effect for camera proximity was 0.05, with a 95% CI of [0.0028, 0.1052], and the proportion mediated was 0.366. The indirect effect for forward lean was 0.04, with a 95% CI of [0.0031, 0.0865], and the proportion mediated was 0.373.

Table 2. Regression Analysis of the Effects of Nonverbal Cues on Immersion.

Variable	Model 1 Baseline	Model 2 Nodding	Model 3 Gesture	Model 4 Movement	Model 5 Proximity	Model 6 Lean	Model 7 Full Model
ln(1 + Nodding)		0.23 * (0.09)					0.21 ** (0.08)
ln(1 + Gesture)			0.27 ** (0.10)				0.25 * (0.10)
ln(1 + Movement)				0.18 * (0.07)			0.16 ** (0.06)
ln(1 + Proximity)					0.17 * (0.07)		0.16 * (0.07)
ln(1 + Lean)						0.14 ** (0.05)	0.14 ** (0.05)
Control variables	YES	YES	YES	YES	YES	YES	YES
Fixed effect (Live session)	YES	YES	YES	YES	YES	YES	YES
Fixed effect (Product category)	YES	YES	YES	YES	YES	YES	YES
VIF	1.25	1.28	1.28	1.28	1.45	1.45	1.45
N	4600	4600	4600	4600	4600	4600	4600
R^2	0.18	0.21	0.23	0.20	0.23	0.22	0.31

Note: ** *p* < 0.01, * *p* < 0.05. Standard errors are in parentheses.

Table 3. Results for the Sales Equation Including Immersion.

Variable	Model 9	Model 10 Immersion
ln(1 + Nodding)	0.17 *** (0.05)	0.17 *** (0.05)
ln(1 + Gesture)	0.16 ** (0.06)	0.14 * (0.06)
ln(1 + Movement)	0.18 ** (0.06)	0.17 ** (0.06)
ln(1 + Proximity)	0.14 *** (0.04)	0.14 *** (0.04)
ln(1 + Lean)	0.12 ** (0.04)	0.13 ** (0.04)
ln(1 + WatchTime)		0.32 ** (0.10)
Control variables	YES	YES
Fixed effect (Live session)	YES	YES
Fixed effect (Product category)	YES	YES
N	4600	4600
R^2	0.39	0.43

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors are in parentheses.

Table 4. Indirect Effects and 95% Confidence Intervals.

Variable	Indirect Effect ($\alpha \times \rho$)	95% CI Lower Bound	95% CI Upper Bound	Proportion Mediated
Nodding frequency	0.07	0.0023	0.1321	0.395
Gesture intensity	0.08	0.0004	0.1596	0.500
Postural movement intensity	0.05	0.0022	0.1002	0.284
Camera proximity	0.05	0.0028	0.1052	0.366
Forward lean	0.04	0.0031	0.0865	0.373

Across all five cues, the confidence intervals for the indirect effects did not include zero, indicating that immersion played a significant mediating role in the effects of the five nonverbal cues on sales. Overall, the results supported immersion as an important process mechanism linking streamers nonverbal presentation to consumer behavior.

5.3. Robustness Check

To assess sensitivity to extreme observations, we re-estimated our main models after excluding segments in the top and bottom 5% of the sales distribution. The results remain qualitatively unchanged: the estimated coefficients on the five nonverbal cues retain the same signs and similar magnitudes, and statistical significance is consistent with the main findings.

6. Conclusions

6.1. Key Findings

This study used merchant-run live sessions on Douyin as the empirical setting. Drawing on social cognitive theory and integrating multimodal video-computing methods, we systematically tested the relationships between streamers' nonverbal cues and consumer purchase behavior and the mediating role of immersion. The results suggested that nodding frequency, gesture intensity, postural movement intensity, camera proximity, and forward lean were all positively associated with consumer purchase behavior, and these findings are consistent with the corresponding main-effect hypotheses.

At the mechanism level, the mediation results indicate that immersion may, to some extent, help explain the pathway through which nonverbal cues are associated with consumer purchase behavior. Specifically, all five nonverbal cues were significantly associated with higher average watch time, and average watch time, in turn, was positively associated with consumer purchase behavior. This result aligns with research on immersion in live streaming, which suggests that immersion is associated with longer viewing duration, lower exit rates, and stronger purchasing tendencies [100]. In contrast, prior mechanism-focused studies in live-streaming commerce have more often attributed the associations of nonverbal cues to increases in credibility, attractiveness, or social presence [11], and they have less frequently tested immersion directly as a key behavioral channel. By introducing average watch time as a process proxy and testing the mediating relationship while controlling for live session and product category differences, this study provided more process-oriented empirical evidence that nonverbal cues are associated with transactions through sustained attention and experiential involvement.

More specifically, the positive association of nodding frequency suggests that streamers' feedback expression during interaction may also be linked to consumer purchase behavior, which is consistent with H1 and H2. The significant pattern for postural movement intensity further suggests that greater overall movement amplitude and more pronounced rhythmic variation may coincide with stronger emotional arousal and a more involving viewing experience, consistent with H3 and H4. The positive association of camera proximity indicates that a closer camera presentation may be linked to viewers' perceptions of distance from the streamer and make it easier for the viewing process to maintain a continuous sense of immersion, in line with H5 and H6. The results for gesture intensity suggest that more dynamic hand movements may be associated with stronger viewers' following of and participation in the live-streaming content, which is consistent with H7 and H8. The significant association of forward lean suggests that a forward-leaning posture may be associated with a stronger sense of proximity and involvement during interaction, making it more likely to be interpreted by viewers as a more active communicative orientation and, in turn, to relate to consumer purchase behavior, consistent with H9 and H10. Overall, although different types of dynamic nonverbal cues may operate in different ways, they all show stable associations with consumer purchase behavior.

This pattern is consistent with prior live-streaming commerce research showing that nonverbal cues can be associated with stronger consumer responses and consumer purchase behavior. For example, studies that used gesture frequency, facial changes, or interaction tempo to explain transaction performance suggested that viewers infer streamers' involvement and communication intent from visible behaviors, which in turn shapes purchase decisions [89]. The difference is that prior work has largely focused on more salient, localized cues such as gaze, facial expressions, and gestures [56]. In contrast, this study extended the scope of nonverbal cues to more processual and rhythmic cues, including head movements, bodily dynamics, and camera proximity. The evidence in this study suggests that these dynamic cues are also significantly associated with consumer purchase behavior, thereby extending the variable coverage of research on nonverbal communication in live-streaming commerce.

6.2. Theoretical Contributions

This study makes three primary theoretical contributions. First, it extends the conceptual boundaries and the set of variables examined in research on nonverbal communication in live-streaming commerce. Existing work has mostly focused on localized cues such as gaze, facial expressions, and gestures [14]. It has rarely examined more processual and manageable presentation features, including head movements, bodily dynamics, and camera proximity, in a systematic way. Using real merchant-run live-session data, we integrated nodding frequency, postural movement intensity, forward lean, and camera proximity into a unified framework and documented their significant associations with consumer purchase behavior. This evidence addressed an empirical gap regarding dynamic behavioral cues in live-streaming commerce research and offered a more comprehensive view of the multidimensional behavioral foundations of live-streaming performance.

Second, this study offers more process-oriented theoretical evidence on the conversion mechanisms of live-streaming commerce. Prior research has often explained the associations of nonverbal cues through relatively static evaluative pathways, such as attractiveness, credibility, or social presence, and has less frequently examined how these cues operate through the viewing process itself [101]. By positioning immersion as the central mediating mechanism, we showed that nonverbal cues are associated with consumer purchase behavior through viewing retention and experiential involvement. This embeds nonverbal presentation into sustained attention and process experience, enriches evidence for applying social cognitive theory in live-streaming commerce, and strengthens the theoretical pathway through which observable behavioral cues are linked to decisions via process experience.

Third, methodologically, this study advances how live-streaming commerce research can leverage unstructured multimodal data. Compared with survey and experimental approaches, which face limitations in ecological validity and scalable measurement, we used AI-based multimodal video analysis to track and quantify nonverbal behaviors at the frame level. We converted behavioral cues into replicable structured measures and aligned them with platform process data at the segment level, enabling large-sample tests in a real transaction environment. This approach enhanced the objectivity and reusability of measurement and provided a useful template for future research that integrates video, operational, and behavioral outcome data in live-streaming settings.

https://doi.org/10.3390/jtaer21040106

6.3. Practical Implications

This study provides actionable managerial implications for live-streaming commerce operations and streamer management. First, the results suggest that improving conversion does not rely solely on more intensive information output. Streamers' on-camera behavioral presentation is also a key lever that can be trained and optimized. Nodding frequency, gesture intensity, postural movement intensity, and forward lean were all positively associated with consumer purchase behavior, and part of these associations was linked to longer average watch time. This implies that streamer training may incorporate on-camera behavior management as a core component. In particular, when responding to chat interactions, emphasizing product benefits and promotion information, and maintaining the pacing of explanations, streamers may use clearer feedback actions and more rhythmic bodily dynamics to sustain viewers' attention and experiential involvement, thereby potentially improving transaction efficiency.

Furthermore, camera proximity, as a highly adjustable media-presentation feature, was significantly associated with consumer purchase behavior and was linked to the immersion mechanism. For merchants and live-room operations, camera distance, framing proportions, and camera strategies across product presentation stages should not be treated merely as stylistic choices. They may be usefully integrated into the conversion-optimization toolbox. Using a closer camera presentation in critical display moments may strengthen psychological proximity and attentional focus, reduce distraction and exit, and ultimately increase viewing continuity and the likelihood of order placement.

Finally, platforms and firms can use nonverbal cue metrics for streamer selection and training evaluation. Because measures such as nodding frequency, gesture intensity, postural movement intensity, and camera proximity can be measured at scale through algorithmic methods and show stable associations with consumer purchase behavior, platforms can incorporate them into streamers' capability profiling and performance review systems. This enables data-driven identification of high-performing behavioral patterns and guidance for novice streamers to develop more effective on-camera expression habits. Compared with experience-based training, this feedback mechanism based on computable behavioral indicators is more likely to yield replicable standards and is better suited for sustained conversion management in the increasingly routine competitive environment of merchant-run live sessions.

6.4. Limitations and Future Research

Although this study used large-scale multimodal data from a real merchant-run live-session setting to test the relationships between streamers' nonverbal cues, consumer purchase behavior, and the immersion mechanism, several limitations remain and warrant further research. First, our identification strategy relied primarily on observational secondary data and fixed-effects controls. As a result, we could not fully rule out time-varying unobserved factors that may jointly affect nonverbal cues and consumer purchase behavior, such as promotion intensity, inventory constraints, pricing strategies, and recommendation order. Future research could strengthen causal identification by leveraging exogenous shocks, quasi-experimental designs, or instrumental-variable approaches. In addition, although this study constructed product-presentation-segment-level data from real live-streaming settings, the final sample may still be subject to certain limitations because live-streaming data are inherently difficult to obtain. For example, in segment-level analysis, multiple observations may correspond to the same streamer or live session. To address this potential issue as far as possible, our models included live session fixed effects and product category fixed effects. At the same time, we implemented a series of rigorous procedures during data processing, including one-to-one matching at the product-presentation-segment level, timestamp alignment, minimum-duration requirements, the exclusion of missing or inconsistent records, and valid-frame coverage screening. In addition, we conducted manual validation of the computer-vision-based measures and evaluated the robustness of the findings through multiple checks, including correlation analysis, multicollinearity diagnostics, and sensitivity analysis excluding extreme observations. The overall pattern of the model estimates, together with related diagnostic evidence, also suggests that the findings are reasonably robust. Nevertheless, future research may further examine this issue by applying additional approaches, such as estimating cluster-robust standard errors clustered at the streamer level, so as to provide a more comprehensive assessment of the robustness of the conclusions.

Second, immersion was proxied by average watch time. While this measure captured viewing retention and sustained attention, it did not cover the full scope of immersion, and watch time may also be driven by information needs or by waiting for benefits such as promotions. In addition, future research could incorporate richer user process participation measures, such as the number of comments, likes, and add-to-cart actions, to more finely capture viewers' interactive participation and different levels of behavioral responses during live-stream viewing. Specifically, comments and likes may, to some extent, reflect viewers' immediate interactive participation and feedback on the content, whereas add-to-cart behavior may more directly capture the transitional stage through which viewing involvement develops toward purchase readiness [102,103].

Third, AI-based measurement of nonverbal cues may still be affected by occlusion, lighting, filters, and camera shake, and segment-level aggregated measures may attenuate the temporal structure of behavior. Future work could improve measurement precision and the explanation of underlying mechanisms by validating algorithmic measures with human annotations and applying temporal modeling to capture dynamic effects at key moments.

Fourth, the sample came from merchant-run live sessions on a single platform, so external validity remains to be established. Future research could extend the setting to more platforms and broader live-streaming contexts, and further examine whether the associations of nonverbal cues vary across different product categories and different live-streaming organizational entities or broadcaster types. Prior research also suggests that product category and broadcaster type may constitute boundary conditions that shape the associations of nonverbal cues [104]. On this basis, future research could further examine boundary conditions such as category involvement, promotion intensity, and streamer type to clarify when nonverbal cues are more effective.

Author Contributions: Conceptualization, X.L. and T.M.; Methodology, X.L. and T.M.; Formal analysis, Q.H. and Q.Y.; Data curation, T.M. and Q.H.; Writing—original draft, X.L., T.M. and Q.H.; Writing—review and editing, Q.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China - Young Scientists Fund Project (72402097), Liaoning Provincial Social Science Youth Project (LJ112410154046), Liaoning University of Technology Doctoral Startup Fund (XB2024012), and Jiangsu Province Graduate Student Practice and Innovation Program (No. SJCX25_1132).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kim, J.; He, N.; Miles, I. Live Commerce Platforms: A New Paradigm for E-Commerce Platform Economy. *J. Theor. Appl. Electron. Commer. Res.* **2023**, *18*, 959–975. [CrossRef]
2. Zhang, C.; Pan, S.; Zhao, Y. More Is Not Always Better: Examining the Drivers of Livestream Sales from an Information Overload Perspective. *J. Retail. Consum. Serv.* **2024**, *77*, 103651. [CrossRef]
3. CNNIC Digital Consumption Development Report (2025)—Internet Development Research. Available online: https://www.cnnic.cn/n4/2025/1226/c88-11445.html (accessed on 19 January 2026).
4. Mutambik, I. The Emerging Phenomenon of Shopstreaming: Gaining a More Nuanced Understanding of the Factors Which Drive It. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 2522–2542. [CrossRef]
5. Xin, M.; Jian, L.; Liu, W.; Bao, Y. Exploring the Effect of Live Streaming Atmospheric Cues on Consumer Impulse Buying: A Flow Experience Perspective. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 149. [CrossRef]
6. Li, N.; Xuan, C.; Chen, R. Different Roles of Two Kinds of Digital Coexistence: The Impact of Social Presence on Consumers' Purchase Intention in the Live Streaming Shopping Context. *J. Retail. Consum. Serv.* **2024**, *80*, 103890. [CrossRef]
7. Tan, J.; Dong, Y.; Zhao, W.; Tan, Q.; Liu, R. Can They Keep You Hooked? Impact of Streamers' Social Capital on User Stickiness in E-Commerce Live Streaming. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 158. [CrossRef]
8. Shi, W.; Li, F.; Hu, M. The Influence of Atmospheric Cues and Social Presence on Consumers' Impulse Buying Behaviors in e-Commerce Live Streaming. *Electron. Commer. Res.* **2025**, *25*, 3325–3353. [CrossRef]
9. Khan, M.; Zeb, S.; Batool, R.; Gasiorowska, A. Non-Verbal Communication Questionnaire: A Measure to Assess Effective Interaction. *Front. Psychol.* **2025**, *15*, 1409675. [CrossRef]
10. Hall, J.A.; Horgan, T.G.; Murphy, N.A. Nonverbal Communication. *Annu. Rev. Psychol.* **2019**, *70*, 271–294. [CrossRef]
11. Chen, Y.; Hu, S.; Wang, L.; Yang, S.; Zhou, S.; Hu, P. The Power of Nonverbal Cues: Evidence from Video Mining in Live Streaming e-Commerce. *J. Retail. Consum. Serv.* **2025**, *87*, 104405. [CrossRef]
12. Liu, L.; Fang, J.; Ji, Z. Posture as Information: Streamer Postures and the Purchase of Micro and Small Enterprise Resellers in Livestreaming. *J. Acad. Mark. Sci.* **2025**, *54*, 254–280. [CrossRef]
13. Miao, H.; Yin, Y.; Zhao, Y.; Gao, Q. Key Elements and Theoretical Logic of Live Streaming E-Commerce Marketing Discourse: An Analysis Based on Grounded Theory. *PLoS ONE* **2025**, *20*, e0322495. [CrossRef] [PubMed]
14. Liu, J.; Zhao, J. Nonverbal Communication of Dual Anchors in Live Streaming and Its Effects on Sales. *J. Retail. Consum. Serv.* **2024**, *81*, 103972. [CrossRef]
15. Hammond, H.; Armstrong, M.; Thomas, G.A.; Gilchrist, I.D. Audience Immersion: Validating Attentional and Physiological Measures against Self-Report. *Cogn. Res.* **2023**, *8*, 22. [CrossRef]
16. Bauer, A.; Kuder, A.; Schulder, M.; Schepens, J. Phonetic Differences between Affirmative and Feedback Head Nods in German Sign Language (DGS): A Pose Estimation Study. *PLoS ONE* **2024**, *19*, e0304040. [CrossRef]
17. Fauviaux, T.; Marin, L.; Parisi, M.; Schmidt, R.; Mostafaoui, G. From Unimodal to Multimodal Dynamics of Verbal and Nonverbal Cues during Unstructured Conversation. *PLoS ONE* **2024**, *19*, e0309831. [CrossRef]
18. Jeitziner, L.T.; Paneth, L.; Rack, O.; Zahn, C. Beyond Words: Investigating Non-Verbal Indicators of Collaborative Engagement in a Virtual Synchronous CSCL Environment. *Front. Psychol.* **2024**, *15*, 1347073. [CrossRef]
19. Li, L.; Zhang, R.; Piper, A.M. Predictors of Student Engagement and Perceived Learning in Emergency Online Education amidst COVID-19: A Community of Inquiry Perspective. *Comput. Hum. Behav. Rep.* **2023**, *12*, 100326. [CrossRef]
20. Arellano-Véliz, N.A.; Cox, R.F.A.; Jeronimus, B.F.; Castillo, R.D.; Kunnen, E.S. Personality Expression in Body Motion Dynamics: An Enactive, Embodied, and Complex Systems Perspective. *J. Res. Pers.* **2024**, *110*, 104495. [CrossRef]
21. Wang, Z.; Luo, C.; Luo, X.; Xu, X. Understanding the Effect of Group Emotions on Consumer Instant Order Cancellation Behavior in Livestreaming E-Commerce: Empirical Evidence from TikTok. *Decis. Support Syst.* **2024**, *179*, 114147. [CrossRef]
22. Fauville, G.; Queiroz, A.C.M.; Luo, M.; Hancock, J.; Bailenson, J.N. Impression Formation from Video Conference Screenshots: The Role of Gaze, Camera Distance, and Angle. *Technol. Mind Behav.* **2022**, *3*, 133–143. [CrossRef]
23. Bandura, A. Toward a Psychology of Human Agency: Pathways and Reflections. *Perspect. Psychol. Sci.* **2018**, *13*, 130–136. [CrossRef] [PubMed]
24. Xin, M.; Liu, W.; Jian, L. Live Streaming Product Display or Social Interaction: How Do They Influence Consumer Intention and Behavior? A Heuristic-Systematic Perspective. *Electron. Commer. Res. Appl.* **2024**, *67*, 101437. [CrossRef]
25. Liu, Z.; Li, J.; Wang, X.; Guo, Y. How Search and Evaluation Cues Influence Consumers' Continuous Watching and Purchase Intentions: An Investigation of Live-Stream Shopping from an Information Foraging Perspective. *J. Bus. Res.* **2023**, *168*, 114233. [CrossRef]
26. Shi, R.; Wang, M.; Qiao, T.; Shang, J. The Effects of Live Streamer's Facial Attractiveness and Product Type on Consumer Purchase Intention: An Exploratory Study with Eye Tracking Technology. *Behav. Sci.* **2024**, *14*, 375. [CrossRef]
27. Yang, Q.; Huo, J.; Li, H.; Xi, Y.; Liu, Y. Can Social Interaction-Oriented Content Trigger Viewers' Purchasing and Gift-Giving Behaviors? Evidence from Live-Streaming Commerce. *Internet Res.* **2023**, *33*, 46–71. [CrossRef]

28. Yang, H.; Wang, B. The Power of Interaction: Fan Growth in Livestreaming E-Commerce. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 203. [CrossRef]
29. Gu, Y.; Cheng, X.; Shen, J. Design Shopping as an Experience: Exploring the Effect of the Live-Streaming Shopping Characteristics on Consumers' Participation Intention and Memorable Experience. *Inf. Manag.* **2023**, *60*, 103810. [CrossRef]
30. Sun, Y.; Shao, X.; Li, X.; Guo, Y.; Nie, K. How Live Streaming Influences Purchase Intentions in Social Commerce: An IT Affordance Perspective. *Electron. Commer. Res. Appl.* **2019**, *37*, 100886. [CrossRef]
31. Li, G.; Jiang, Y.; Chang, L. The Influence Mechanism of Interaction Quality in Live Streaming Shopping on Consumers' Impulsive Purchase Intention. *Front. Psychol.* **2022**, *13*, 918196. [CrossRef]
32. Park, H.J.; Lin, L.M. The Effects of Match-Ups on the Consumer Attitudes toward Internet Celebrities and Their Live Streaming Contents in the Context of Product Endorsement. *J. Retail. Consum. Serv.* **2020**, *52*, 101934. [CrossRef]
33. Hu, M.; Zhang, M.; Wang, Y. Why Do Audiences Choose to Keep Watching on Live Video Streaming Platforms? An Explanation of Dual Identification Framework. *Comput. Hum. Behav.* **2017**, *75*, 594–606. [CrossRef]
34. Carmichael, C.L.; Mizrahi, M. Connecting Cues: The Role of Nonverbal Cues in Perceived Responsiveness. *Curr. Opin. Psychol.* **2023**, *53*, 101663. [CrossRef] [PubMed]
35. Patterson, M.L.; Fridlund, A.J.; Crivelli, C. Four Misconceptions About Nonverbal Communication. *Perspect. Psychol. Sci.* **2023**, *18*, 1388–1411. [CrossRef] [PubMed]
36. Van Zant, A.B.; Berger, J. How the Voice Persuades. *J. Pers. Soc. Psychol.* **2020**, *118*, 661–682. [CrossRef]
37. Burgoon, J.K.; Wang, X.; Chen, X.; Pentland, S.J.; Dunbar, N.E. Nonverbal Behaviors "Speak" Relational Messages of Dominance, Trust, and Composure. *Front. Psychol.* **2021**, *12*, 624177. [CrossRef]
38. Cascio Rizzo, G.L.; Berger, J.; Zhou, M. Talking with Your Hands: How Hand Gestures Influence Communication. *J. Mark. Res.* **2025**. [CrossRef]
39. Jarick, M.; Bencic, R. Eye Contact Is a Two-Way Street: Arousal Is Elicited by the Sending and Receiving of Eye Gaze Information. *Front. Psychol.* **2019**, *10*, 1262. [CrossRef]
40. Todorov, A.; Olivola, C.Y.; Dotsch, R.; Mende-Siedlecki, P. Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annu. Rev. Psychol.* **2015**, *66*, 519–545. [CrossRef]
41. Zhang, S.; Friedman, E.M.S.; Srinivasan, K.; Dhar, R.; Zhang, X. Serving with a Smile on Airbnb: Analyzing the Economic Returns and Behavioral Underpinnings of the Host's Smile. *J. Consum. Res.* **2025**, *51*, 1073–1097. [CrossRef]
42. Höfling, T.T.A.; Alpers, G.W. Automatic Facial Coding Predicts Self-Report of Emotion, Advertisement and Brand Effects Elicited by Video Commercials. *Front. Neurosci.* **2023**, *17*, 1125983. [CrossRef] [PubMed]
43. Liu, X.; Shi, S.W.; Teixeira, T.; Wedel, M. Video Content Marketing: The Making of Clips. *J. Mark.* **2018**, *82*, 86–101. [CrossRef]
44. Kim, W.; Sung, J.; Saakes, D.; Huang, C.; Xiong, S. Ergonomic Postural Assessment Using a New Open-Source Human Pose Estimation Technology (OpenPose). *Int. J. Ind. Ergon.* **2021**, *84*, 103164. [CrossRef]
45. Kujani, T.; Kumar, V.D. Head Movements for Behavior Recognition from Real Time Video Based on Deep Learning ConvNet Transfer Learning. *J. Ambient Intell. Humaniz. Comput.* **2023**, *14*, 7047–7061. [CrossRef]
46. Li, M.; Hua, Y. Integrating Social Presence with Social Learning to Promote Purchase Intention: Based on Social Cognitive Theory. *Front. Psychol.* **2022**, *12*, 810181. [CrossRef]
47. Schunk, D.H.; DiBenedetto, M.K. Motivation and Social Cognitive Theory. *Contemp. Educ. Psychol.* **2020**, *60*, 101832. [CrossRef]
48. Lo Schiavo, M.; Prinari, B.; Saito, I.; Shoji, K.; Benight, C.C. A Dynamical Systems Approach to Triadic Reciprocal Determinism of Social Cognitive Theory. *Math. Comput. Simul.* **2019**, *159*, 18–38. [CrossRef]
49. Hwang, J.; Youn, S. From Brick-and-Mortar to Livestream Shopping: Product Information Acquisition from the Uncertainty Reduction Perspective. *Fash. Text.* **2023**, *10*, 7. [CrossRef]
50. Lim, X.-J.; Luo, X.; Cheah, J.-H.; Tan, K.-L.; Hall, C.M. Unveiling Impulse Buying Patterns in Travel Live-Streaming through the Lens of Social Cognitive Theory. *J. Vacat. Mark.* **2026**, *32*, 19–34. [CrossRef]
51. Lim, J.S.; Choe, M.-J.; Zhang, J.; Noh, G.-Y. The Role of Wishful Identification, Emotional Engagement, and Parasocial Relationships in Repeated Viewing of Live-Streaming Games: A Social Cognitive Theory Perspective. *Comput. Hum. Behav.* **2020**, *108*, 106327. [CrossRef]
52. Tian, Y.; Frank, B. Optimizing Live Streaming Features to Enhance Customer Immersion and Engagement: A Comparative Study of Live Streaming Genres in China. *J. Retail. Consum. Serv.* **2024**, *81*, 103974. [CrossRef]
53. Tang, X.; Hao, Z.; Li, X. The Influence of Streamers' Physical Attractiveness on Consumer Response Behavior: Based on Eye-Tracking Experiments. *Front. Psychol.* **2024**, *14*, 1297369. [CrossRef] [PubMed]
54. Strauss, D.J.; Francis, A.L.; Vibell, J.; Corona–Strauss, F.I. The Role of Attention in Immersion: The Two–Competitor Model. *Brain Res. Bull.* **2024**, *210*, 110923. [CrossRef] [PubMed]
55. Joo, E.; Yang, J. How Perceived Interactivity Affects Consumers' Shopping Intentions in Live Stream Commerce: Roles of Immersion, User Gratification and Product Involvement. *J. Res. Interact. Mark.* **2023**, *17*, 754–772. [CrossRef]

56. Zhang, Y.; Li, X.; Liu, R.; Shuai, Q.; Huang, C. How Facial Cues of Streamers Drive Purchase Intention in Live Streaming Commerce: Based on a Lens Model. *J. Bus. Res.* **2025**, *190*, 115193. [CrossRef]
57. Xian, T.; Fang, Z.; Huang, Z.; Niu, Y.; Liao, C. Benefit Appeals in Livestreaming E-Commerce: The Moderating Role of Influencers' Facial Beauty and Body Motion. *J. Advert. Res.* **2025**, 1–33. [CrossRef]
58. Pang, H.T.; Zhou, X.; Chu, M. Cross-Cultural Differences in Using Nonverbal Behaviors to Identify Indirect Replies. *J. Nonverbal Behav.* **2024**, *48*, 323–344. [CrossRef]
59. Wang, Y.; Lu, Z.; Cao, P.; Chu, J.; Wang, H.; Wattenhofer, R. How Live Streaming Changes Shopping Decisions in E-Commerce: A Study of Live Streaming Commerce. *Comput. Support. Coop. Work* **2022**, *31*, 701–729. [CrossRef]
60. Aburumman, N.; Gillies, M.; Ward, J.A.; Hamilton, A.F.d.C. Nonverbal Communication in Virtual Reality: Nodding as a Social Signal in Virtual Interactions. *Int. J. Hum.-Comput. Stud.* **2022**, *164*, 102819. [CrossRef]
61. Osugi, T.; Kawahara, J.I. Effects of Head Nodding and Shaking Motions on Perceptions of Likeability and Approachability. *Perception* **2018**, *47*, 16–29. [CrossRef]
62. Wang, K.; Zhai, X.; Cheung, M.K.M.; Fu, E.Y.; Chen, P.Q.; Ngai, G.; Leong, H.V. TMAN: A Temporal Multimodal Attention Network for Backchannel Detection. *Neurocomputing* **2025**, *657*, 131605. [CrossRef]
63. Dayan, E.; Barliya, A.; de Gelder, B.; Hendler, T.; Malach, R.; Flash, T. Motion Cues Modulate Responses to Emotion in Movies. *Sci. Rep.* **2018**, *8*, 10881. [CrossRef] [PubMed]
64. Williams, E.H.; Cristino, F.; Cross, E.S. Human Body Motion Captures Visual Attention and Elicits Pupillary Dilation. *Cognition* **2019**, *193*, 104029. [CrossRef] [PubMed]
65. Su, Q.; Li, F. Influence of Time Metaphor and Destination Image Proximity on Tourist Responses. *Tour. Manag.* **2024**, *105*, 104942. [CrossRef]
66. Kim, K.; Lee, S.; Choi, Y.K. Image Proximity in Advertising Appeals: Spatial Distance and Product Types. *J. Bus. Res.* **2019**, *99*, 490–497. [CrossRef]
67. Veranic, K.; Ewing, L.; Sambrook, T.; Watson, E.A.G.; Zhao, M.; Bayliss, A.P. Changes in Interpersonal Distance Modulate Social Attention Engagement: Evidence from EEG Alpha Band Suppression. *Soc. Cogn. Affect. Neurosci.* **2025**, *20*, nsaf008. [CrossRef]
68. Liao, J.; Chen, K.; Qi, J.; Li, J.; Yu, I.Y. Creating Immersive and Parasocial Live Shopping Experience for Viewers: The Role of Streamers' Interactional Communication Style. *J. Res. Interact. Mark.* **2023**, *17*, 140–155. [CrossRef]
69. Rodero, E. Effectiveness, Attractiveness, and Emotional Response to Voice Pitch and Hand Gestures in Public Speaking. *Front. Commun.* **2022**, *7*, 869084. [CrossRef]
70. Zhang, G.; Cao, J.; Liu, D. Examining the Influence of Information Overload on Consumers' Purchase in Live Streaming: A Heuristic-Systematic Model Perspective. *PLoS ONE* **2023**, *18*, e0284466. [CrossRef]
71. Bujok, R.; Peeters, D.; Meyer, A.S.; Bosker, H.R. Beating Stress: Evidence for Recalibration of Word Stress Perception. *Atten. Percept. Psychophys.* **2025**, *87*, 1729–1749. [CrossRef]
72. Ferrari, A.; Hagoort, P. Beat Gestures and Prosodic Prominence Interactively Influence Language Comprehension. *Cognition* **2025**, *256*, 106049. [CrossRef]
73. Sauter, D.A. The Nonverbal Communication of Positive Emotions: An Emotion Family Approach. *Emot. Rev.* **2017**, *9*, 222–234. [CrossRef] [PubMed]
74. Guo, J.; Li, Y.; Xu, Y.; Zeng, K. How Live Streaming Features Impact Consumers' Purchase Intention in the Context of Cross-Border E-Commerce? A Research Based on SOR Theory. *Front. Psychol.* **2021**, *12*, 767876. [CrossRef] [PubMed]
75. Shi, Y.; Ma, C.; Zhu, Y. The Impact of Emotional Labor on User Stickiness in the Context of Livestreaming Service—Evidence from China. *Front. Psychol.* **2021**, *12*, 698510. [CrossRef] [PubMed]
76. Liu, H.; Liang, J. A Study on the Factors Influencing Chinese Costume Consumers Utilizing Live Streaming Platforms to Purchase Products: A Case Study of Douyin. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 38. [CrossRef]
77. Wang, X.; Tian, Y.; Lan, R.; Yang, W.; Zhang, X. Beyond the Watching: Understanding Viewer Interactions in Crowdsourced Live Video Broadcasting Services. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *29*, 3454–3468. [CrossRef]
78. Lugaresi, C.; Tang, J.; Nash, H.; McClanahan, C.; Uboweja, E.; Hays, M.; Zhang, F.; Chang, C.-L.; Yong, M.G.; Lee, J.; et al. MediaPipe: A Framework for Building Perception Pipelines. *arXiv* **2019**, arXiv:1906.08172. [CrossRef]
79. Zhang, F.; Bazarevsky, V.; Vakunov, A.; Tkachenka, A.; Sung, G.; Chang, C.-L.; Grundmann, M. MediaPipe Hands: On-Device Real-Time Hand Tracking. *arXiv* **2020**, arXiv:2006.10214.
80. Ramseyer, F.T. Motion Energy Analysis (MEA): A Primer on the Assessment of Motion from Video. *J. Couns. Psychol.* **2020**, *67*, 536–549. [CrossRef]
81. Tharatipyakul, A.; Srikaewsiew, T.; Pongnumkul, S. Deep Learning-Based Human Body Pose Estimation in Providing Feedback for Physical Movement: A Review. *Heliyon* **2024**, *10*, e36589. [CrossRef]
82. Xiang, A.; Andrews, J.T.A.; Bourke, R.L.; Thong, W.; LaChance, J.M.; Georgievski, T.; Modas, A.; Rahmattalabbi, A.; Ba, Y.; Nagpal, S.; et al. Fair Human-Centric Image Dataset for Ethical AI Benchmarking. *Nature* **2025**, *648*, 97–108. [CrossRef] [PubMed]

83. Hammadi, Y.; Grondin, F.; Ferland, F.; Lebel, K. Evaluation of Various State of the Art Head Pose Estimation Algorithms for Clinical Scenarios. *Sensors* **2022**, *22*, 6850. [CrossRef] [PubMed]
84. Iijima, S.; Shiomi, M.; Hara, T. Verification of Reliability and Validity of Trunk Forward Tilt Angle Measurement During Gait Using 2-Dimensional Motion Analysis. *J. Chiropr. Med.* **2023**, *22*, 89–95. [CrossRef] [PubMed]
85. Zheng, S.; Chen, J.; Liao, J.; Hu, H.-L. What Motivates Users' Viewing and Purchasing Behavior Motivations in Live Streaming: A Stream-Streamer-Viewer Perspective. *J. Retail. Consum. Serv.* **2023**, *72*, 103240. [CrossRef]
86. Barta, S.; Gurrea, R.; Flavián, C. Telepresence in Live-Stream Shopping: An Experimental Study Comparing Instagram and the Metaverse. *Electron. Mark.* **2023**, *33*, 29. [CrossRef]
87. He, L.; Li, X.; Li, Y.; Liu, Y.; Zhang, N.; Zhou, X. Is More Always Better? The Effect of Audience Size on Sales Performance in Live Streaming Commerce: A Multimethod Study. *J. Retail. Consum. Serv.* **2026**, *89*, 104613. [CrossRef]
88. Yang, G.; Chaiyasoonthorn, W.; Chaveesuk, S. Exploring the Influence of Live Streaming on Consumer Purchase Intention: A Structural Equation Modeling Approach in the Chinese E-Commerce Sector. *Acta Psychol.* **2024**, *249*, 104415. [CrossRef]
89. Zou, J.; Fu, X. Understanding the Purchase Intention in Live Streaming from the Perspective of Social Image. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 1500. [CrossRef]
90. Liu, D.; Yu, J. Impact of Perceived Diagnosticity on Live Streams and Consumer Purchase Intention: Streamer Type, Product Type, and Brand Awareness as Moderators. *Inf. Technol. Manag.* **2024**, *25*, 219–232. [CrossRef]
91. Li, M.; Cheng, M.; Quintal, V. Decoding the Subtleties: Speech Voice Cues and Their Impacts on Viewer In-Consumption Engagement in Travel Live Streaming. *J. Hosp. Tour. Res.* **2025**. [CrossRef]
92. Luo, L.; Xu, M.; Wan, F. Is Sadness Always Bad? The Effects of Streamers' Multimodal Expressions on Viewer Engagement in Live Streaming. *J. Retail. Consum. Serv.* **2026**, *89*, 104579. [CrossRef]
93. Yang, Q.; Wang, Y.; Song, M.; Jiang, Y.; Li, Q. Sonic Strategies: Unveiling the Impact of Sound Features in Short Video Ads on Enterprise Market Entry Performance. *J. Bus.Bus. Mark.* **2025**, *32*, 95–116. [CrossRef]
94. Miao, M.; Wang, Y.; Li, J.; Jiang, Y.; Yang, Q. Audio Features and Crowdfunding Success: An Empirical Study Using Audio Mining. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 3176–3196. [CrossRef]
95. Dover, Y.; Amichai-Hamburger, Y. Characteristics of Online User-Generated Text Predict the Emotional Intelligence of Individuals. *Sci. Rep.* **2023**, *13*, 6778. [CrossRef]
96. Markowitz, D.M.; Shulman, H.C. The Predictive Utility of Word Familiarity for Online Engagements and Funding. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2026045118. [CrossRef]
97. Schoenmueller, V.; Netzer, O.; Stahl, F. The Polarity of Online Reviews: Prevalence, Drivers and Implications. *J. Mark. Res.* **2020**, *57*, 853–877. [CrossRef]
98. Lachowicz, M.J.; Preacher, K.J.; Kelley, K. A Novel Measure of Effect Size for Mediation Analysis. *Psychol. Methods* **2018**, *23*, 244–261. [CrossRef]
99. Kim, J.H. Multicollinearity and Misleading Statistical Results. *Korean J. Anesthesiol.* **2019**, *72*, 558–569. [CrossRef]
100. Jiang, Y.; Lee, H.-T.; Li, W. The Effects of Live Streamer's Expertise and Entertainment on the Viewers' Purchase and Follow Intentions. *Front. Psychol.* **2024**, *15*, 1383736. [CrossRef]
101. Song, S.; Xu, Y.; Ma, B.; Zong, X. Understanding How Streamer's Self-Presentation in E-Commerce Live Streaming Affects Consumers: The Role of Persuasion Knowledge. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 1922–1942. [CrossRef]
102. Niu, W.; Zhang, W.; Chen, M.; Han, M. Prefer Cheap or Expensive Products? Shopping Stage Matters. *Front. Psychol.* **2024**, *15*, 1418082. [CrossRef]
103. Thai, T.D.-H.; Wang, T. Investigating the Effect of Social Endorsement on Customer Brand Relationships by Using Statistical Analysis and Fuzzy Set Qualitative Comparative Analysis (fsQCA). *Comput. Hum. Behav.* **2020**, *113*, 106499. [CrossRef]
104. Yang, X.; Liu, Y.; Dong, J.; Li, S. Impact of Streamers' Characteristics on Sales Performance of Search and Experience Products: Evidence from Douyin. *J. Retail. Consum. Serv.* **2023**, *70*, 103155. [CrossRef]