



## A MULTIMODAL DISCOURSE ANALYSIS OF GESTURES AND BODY LANGUAGE IN CONSTRUCTING SUFI IDENTITY IN PAKISTANI MUSIC VIDEOS

**Hafiza Rumaisa Rehman Rao**

Rumaisa.rehman02@gamil.com

PhD Scholar, Department of English, University of Education, Lahore, Pakistan

**Eisha Tur Razia** (corresponding author)

eishaturrazia999@gmail.com

PhD Scholar, Department of English, University of Education, Lahore, Pakistan

**Dr. Humaira Irfan**

Associate Professor of English, Department of English, University of Education, Lahore, Pakistan.

### Abstract

*This study presents a qualitative multimodal discourse analysis of gesture and body language in three contemporary Pakistani Sufi music videos: Sultan Ateeq Rehman's Asaa Ik Din Sab Nu Yad Araan, Ramzan Jani's Aj Aakhan Waris Shah Nu, and Muhammad Samie's Maikada 3: Bus Ik Nigah. Adopting a case study design, the analysis integrates McNeill's (1992) gesture typology, Kress and van Leeuwen's (2020) social semiotic visual grammar, and Halid's (2022) framework of spiritual performativity to examine how gesture and body language function as semiotic resources in constructing Sufi identity within the digital media context. Systematic gesture coding and multimodal discourse interpretation reveal a shared and culturally specific gestural vocabulary across all three performers, encompassing the upward gaze encoding hal (spiritual absorption), the bilateral open-palm gesture encoding dua (supplication), the deictic arm extension encoding munajat (petitionary invocation), and the raised index finger encoding tawhid (divine unity), alongside significant performer-specific variation that constructs distinct Sufi subject positions. The analysis identifies three principal mechanisms through which gesture encodes Sufi theological concepts: spatial metaphor organised along the body's vertical axis, semiotic synchrony between gesture strokes and key Sufi vocabulary items, and performative embodiment whereby the body enacts rather than represents spiritual states. The findings establish that gesture in Sufi music video is constitutive of theological meaning rather than supplementary to it, and that the performer's body functions as the primary site through which centuries-old devotional traditions are sustained and communicated to contemporary digital audiences. The study advances a replicable methodological model for applying gesture analysis to mediated religious performance in non-Western contexts, contributing to multimodal discourse studies, Pakistani media research, and the broader study of how spiritual traditions maintain coherence through algorithmically mediated content.*

**Keywords:** *Multimodal Discourse Analysis, Gesture, Pakistani Music Videos, Body Language, Social Semiotics, Sufi Identity*

### 1. Introduction

Sufi music has occupied a central place in the devotional and cultural life of Pakistan since the country's independence, drawing on centuries of poetic and musical tradition associated with figures such as Bulleh Shah, Waris Shah, and Mian Muhammad Bakhsh (Mir, 2010). In the contemporary digital era, this tradition has migrated from shrines, dargahs, and private mehfilis onto YouTube and social media platforms, where individual performers produce professionally filmed music videos that circulate to audiences numbering in the millions (Bunt, 2018; Williams & Mahmood, 2019). This transformation raises a question that is simultaneously cultural and linguistic: how, in the absence of a live devotional gathering, does a Sufi music video communicate the spiritual identity and theological content that the tradition carries? This study proposes that the answer lies substantially in the body. Gestures, gaze, posture, and bodily orientation are not



incidental to the performance of Sufi music; they are among its primary meaning-making resources, translating the abstract theological vocabulary of the kalam into visible, embodied signs that audiences can read and respond to.

This study analyses three contemporary Pakistani Sufi music videos through a multimodal discourse framework, examining how gesture and body language construct a recognisable Sufi identity across different performers, linguistic registers, and visual settings. The videos are Sultan Ateeq Rehman's *Asaa Ik Din Sab Nu Yad Araan*, a rendition of classical Punjabi kalam by the 19th-century Kashmiri Sufi poet Mian Muhammad Bakhsh; Ramzan Jani's *Aj Aakhan Waris Shah Nu*, a performance of Amrita Pritam's 1947 partition poem addressed to the Sufi saint Waris Shah; and Muhammad Samie's *Maikada 3: Bus Ik Nigah*, an original Urdu composition drawing on classical Sufi wine-metaphor imagery.

### 1.1 Background

The study of Sufi music in Pakistan has a well-established scholarly foundation. Qureshi's (1986) seminal ethnomusicological work on qawwali documented the performance conventions through which Sufi music induces spiritual states in its audience, while Sakata (1994) analysed the performances of Nusrat Fateh Ali Khan as instances of embodied theological communication. More recently, Williams and Mahmood (2019) have examined how digital media platforms have transformed the production and reception of Pakistani music, arguing that music videos have become significant sites for the construction and negotiation of cultural and religious identity. What remains underexplored in this body of scholarship is the specific role of gesture and body language as semiotic resources in Sufi music video performance.

Multimodal discourse analysis, developed principally by Kress and van Leeuwen (2020), provides the tools to address this gap. Their social semiotic framework treats all modes of communication, including gesture, spatial arrangement, gaze, and visual composition, as meaning-making resources operating within culturally specific codes. Fairclough's (2013) critical discourse analysis extends this to examine how such modes construct ideological positions, identities, and power relations. Applied to Sufi music videos, this framework enables systematic analysis of how bodily and visual modes work together to produce the genre's distinctive spiritual meanings.

Gesture theory contributes a further layer of analytical precision. McNeill (1992) demonstrated that gesture is not supplementary to speech but co-constitutive of meaning, sharing a cognitive origin with the verbal content it accompanies. His typology of iconic, metaphoric, deictic, and beat gestures provides a principled vocabulary for coding the movements observed in the videos. Kendon's (2004) concept of gesture as visible action as utterance reinforces the status of bodily movement as a form of communication in its own right, not merely an illustration of verbal content.

### 1.2 Problem Statement and Rationale

While the scholarship reviewed above establishes robust foundations for the present study, a clear analytical gap remains. No prior research has systematically applied a multimodal gesture analysis framework to contemporary Pakistani Sufi music videos. Existing studies have examined either the musical and social dimensions of live Sufi performance (Qureshi, 1986; Sakata, 1994) or the broader cultural and political dimensions of Pakistani music media (Williams & Mahmood, 2019; Sarrazin, 2013), but none has asked how specific gesture types encode Sufi theological concepts across multiple performers in the digital music video format. As Sufi music videos reach increasingly global audiences through YouTube, understanding the semiotic mechanisms through



which they construct and communicate Sufi identity becomes both a scholarly and a cultural necessity. This study addresses that gap by combining gesture typology, visual grammar, and discourse analysis in a systematic investigation of three representative videos.

#### **1.4 Research Objectives**

This study aims to:

1. Identify what gestures and body language appear across contemporary Pakistani Sufi music videos.
2. Examine how those gestures and body language function to convey key Sufi concepts in these music videos.

#### **1.5 Research Questions**

The study is guided by two research questions:

1. What gestures and body language are used in contemporary Pakistani Sufi music videos?
2. How do these specific gestures and body language encode key Sufi concepts across these music videos?

#### **1.6 Significance of the Study**

This study contributes to three intersecting fields. For multimodal discourse studies, it provides a methodological model for applying gesture analysis to mediated religious performance in non-Western contexts. For Pakistani media studies, it offers the first systematic multimodal analysis of spiritual gesture in contemporary Sufi music videos, filling an empirical gap identified in the literature. For Sufi studies more broadly, it demonstrates how a centuries-old tradition of embodied devotional practice is being adapted and transmitted through the visual grammar of digital media, sustaining theological meanings for audiences who may never attend a live performance. The study is also of practical significance for researchers and practitioners in religious communication, cultural heritage, and digital media studies who are concerned with how spiritual traditions maintain their identity and coherence in the age of algorithmically mediated content.

### **3. Literature Review**

Sufi music performance is a site where embodied meaning, spiritual identity, and religious tradition converge, and understanding it demands frameworks that account for the simultaneous operation of multiple semiotic resources. This review examines the theoretical and empirical foundations that underpin the current study, drawing on scholarship in multimodal discourse analysis, gesture studies, embodied performance, and Pakistani Sufi music culture.

The analytical framework for this study is grounded in Kress and van Leeuwen's social semiotic theory of multimodality, which holds that meaning is distributed across multiple modes rather than residing in language alone (Kress & van Leeuwen, 2020). Their visual grammar provides tools for analysing how compositional choices, framing, and salience work together to produce meaning in image-based texts. Van Leeuwen (2023) further develops the principle that each mode carries culturally shaped affordances, enabling and constraining the kinds of meaning that can be made within it. For music video analysis, this means that camera angle, spatial arrangement, performer positioning, and the editing relationship between gesture and lyric are all meaning-making resources that must be read together rather than in isolation.

Multimodal Critical Discourse Analysis, as developed by Fairclough (2013) and applied by recent scholars in Pakistani contexts (Jabbar et al., 2025; Qayyum & Ali, 2024; Taj & Ameen,



2024), extends these semiotic insights to examine how power, ideology, and cultural identity are constructed through combined verbal, visual, and gestural modes. In religious media specifically, such analysis can uncover how spiritual authority is visually encoded and how bodies become sites of ideological positioning. The present study applies MCDA to Sufi music videos to examine how gesture and body language construct a distinctly Sufi identity for Pakistani audiences.

Gesture theory establishes that bodily movement is not incidental to communication but constitutive of it. McNeill (1992) demonstrated that gesture and speech form an integrated cognitive and communicative system, where gesture carries meanings that either complement or extend verbal content. His typology distinguishes iconic gestures, which visually depict referential content; metaphoric gestures, which give spatial form to abstract ideas; deictic gestures, which point toward real or imagined locations; and beat gestures, which mark prosodic structure. Kendon (2004) further theorised gesture as visible action as utterance, emphasising its intentional communicative status across cultural contexts. These frameworks are central to the present study's method of coding and interpreting the gestures observed in three Pakistani Sufi music videos.

The relationship between gesture and spiritual content has received attention in performance studies. Halid (2022) analyses embodied practices in Sufi contexts, arguing that physical movements in devotional performance are not representational but constitutive: performers do not merely depict spiritual states but enact and transmit them through the body. This understanding shapes the interpretive logic of the current analysis, where each coded gesture is treated not as decoration but as a primary vehicle for spiritual meaning.

The scholarship on South Asian Sufi music provides essential contextual grounding. Qureshi's (1986) foundational study of qawwali documents how performers use vocal technique, rhythmic patterning, and bodily movement to induce states of spiritual awareness in audiences, a process she terms response behaviour. Sakata (1994) extends this to the performances of Nusrat Fateh Ali Khan, demonstrating how a single performer's gesture vocabulary can communicate theological concepts and simultaneously signal the performer's own spiritual state. These studies establish that in the Pakistani Sufi tradition, the body carries theological weight: posture, gaze, and hand movement are not stylistic choices but semantic ones.

Sarrazin (2013) examines how qawwali is mediated in visual formats, noting that the camera's framing decisions transform devotional performance into a constructed visual argument. Williams and Mahmood (2019) trace the evolution of Pakistani music media from radio to digital platforms, showing how Coke Studio and YouTube have changed how Sufi content is produced and consumed. Contemporary Sufi music videos are not simply recordings of live performance but crafted multimodal texts in which visual production choices extend and sometimes reframe the spiritual meanings of the sung kalam (sacred poetry). This study builds on these observations by systematically analysing the gestural dimension of such videos.

Despite this body of scholarship, a clear gap exists at the intersection of gesture analysis and Pakistani Sufi music video research. Existing studies have examined either the musical and sociological dimensions of Sufi performance in live contexts (Qureshi, 1986; Sakata, 1994) or the broader visual and cultural dimensions of Pakistani music media (Williams & Mahmood, 2019; Sarrazin, 2013), but none has applied a systematic gesture analysis framework to contemporary Pakistani Sufi music videos. McNeill's (1992) and Kendon's (2004) typologies have been used in conversational and face-to-face interaction research but have not been applied to mediated Sufi performance, where gesture is shaped by both traditional devotional conventions and the visual-



compositional choices of music video production. Similarly, while MCDA has been applied to Pakistani media (Jabbar et al., 2025; Taj & Ameen, 2024), religious and spiritual content in music videos remains largely unexamined within this framework. No prior study has investigated how specific gesture types encode Sufi theological concepts across multiple contemporary Pakistani music videos, nor has any work traced how body language constructs a coherent Sufi identity in this genre. This study addresses that gap directly, providing both a methodological model and empirical data for this underexplored area.

### **3. Methodology**

#### **3.1 Research Design**

This study adopts a qualitative research design, which is appropriate given its aim of producing a detailed, interpretive account of how gesture and body language construct meaning in a specific cultural and media context. Qualitative research is well suited to questions of meaning-making and identity construction, as it allows for the kind of close, contextually sensitive analysis that such questions require (Denzin & Lincoln, 2018). A case study approach structures the investigation: each of the three selected music videos constitutes a bounded case, and the analysis moves from individual case analysis to cross-case comparison. This design enables the identification of both video-specific gestural features and shared patterns that characterise the genre as a whole.

The epistemological basis of the study is interpretive, meaning that the researcher treats knowledge as constructed through the interaction between the analyst's theoretical framework and the data rather than as an objective property of the texts themselves (Creswell & Creswell, 2018). This position is consistent with the social semiotic tradition, in which meaning is understood as culturally and contextually produced rather than inherent in signs. The analytical process is therefore both systematic and interpretive: systematic in its application of a defined coding framework drawn from gesture theory and visual grammar, and interpretive in its reading of how coded elements combine to produce culturally specific meanings within the Sufi performance tradition.

#### **3.2 Research Method**

The primary method is multimodal discourse analysis (MDA), combining Kress and van Leeuwen's (2020) visual grammar with McNeill's (1992) gesture typology. MDA is a systematic analytical method for examining how multiple communicative modes, including language, image, gesture, and spatial arrangement, work together to produce meaning in a text. In the context of music video analysis, this means examining simultaneously the verbal content of the kalam, the visual composition of the frame, and the gestures and postural configurations of the performer, treating each as a semiotic resource that contributes to the overall meaning of the performance.

The analytical procedure follows three stages. First, each video is viewed repeatedly in its entirety to develop familiarity with its structure, content, and visual style. Second, a systematic gesture inventory is compiled for each video, coding each identifiable gesture according to McNeill's (1992) typology: iconic gestures that visually depict referential content; metaphoric gestures that give spatial form to abstract concepts; deictic gestures that point toward real or imagined locations; and beat gestures that mark prosodic rhythm. Third, each coded gesture is interpreted in relation to the verbal anchor with which it co-occurs, drawing on Kress and van Leeuwen's (2020) compositional analysis to examine how camera framing, shot scale, and visual



saliency shape the gesture's meaning within the video text. The Sufi theological concepts encoded by each gesture are identified through reference to established scholarship on Sufi devotional practice (Qureshi, 1986; Sakata, 1994; Halid, 2022).

### **3.3 Data Collection and Sampling**

Data were collected purposively from YouTube, selecting the most-viewed video of each of three contemporary Pakistani Sufi singers. Purposive sampling is appropriate when the aim is to select information-rich cases rather than a statistically representative sample (Creswell & Creswell, 2018). The selection criteria were: all performers are contemporary Pakistani artists working within the Sufi kalam tradition; each video is the most-viewed upload on the respective singer's channel, indicating wide cultural circulation; and the three videos together represent distinct linguistic registers, namely classical Pothohari Punjabi, literary Punjabi, and classical Urdu, enabling comparison across language variation within the genre.

The three videos are Sultan Ateeq Rehman's *Asaa Ik Din Sab Nu Yad Araan* (Kalam Mian Muhammad Bakhsh Part 2, approximately 5 minutes 46 seconds), Ramzan Jani's *Aj Aakhan Waris Shah Nu* (Sufi Kalam, 2024, approximately 4 minutes 50 seconds), and Muhammad Samie's *Maikada 3: Bus Ik Nigah* (Official Video, approximately 14 minutes 36 seconds). Data collected from each video consisted of timestamped frames capturing key gestural moments, supplemented by close listening to identify the verbal anchors with which gestures co-occur. Lyric subtitles visible in two of the three videos supported the verbal analysis.

### **3.4 Theoretical Framework**

The three theoretical frameworks, used in combination, guide the analysis. Kress and van Leeuwen's (2021) social semiotic multimodality framework provide the overarching analytical vocabulary, treating gesture, gaze, posture, spatial arrangement, and visual composition as culturally shaped meaning-making modes that must be read together rather than in isolation. Their compositional grammar informs the analysis of camera angles, shot types, and performer positioning, treating the Sufi music video as a constructed multimodal text in which every visual and bodily element carries semantic weight.

McNeill's (1992) gesture theory provides the primary tool for coding and interpreting performers' physical movements. His typology of iconic, metaphoric, deictic, and beat gestures offers a principled vocabulary for classifying the movements observed across the three videos, while his concept of the growth point, the cognitive unit from which gesture and speech are co-generated, informs the study's attention to synchrony between gestural strokes and key Sufi vocabulary items.

Halid's (2022) framework of Sufi spiritual performativity supplies the theologically grounded interpretive layer. Drawing on his analysis of enduring Sufi devotional traditions, Halid argues that bodily practices in Sufi contexts are constitutive rather than representational: the body does not illustrate a spiritual state but enacts and transmits it. This framework enables the analysis to move beyond identifying what gestures mean to explaining why they carry devotional authority, grounding interpretation in the specific theological logic of the tradition under study.

Together, these frameworks form a mutually reinforcing analytical architecture in which Kress and van Leeuwen (2021) account for visual and compositional meaning, McNeill (1992) accounts for gesture mechanics, and Halid (2022) accounts for the spiritual performative logic that makes specific bodily forms theologically significant within the Sufi tradition.

### **3.5 Ethical Considerations**

All three videos are publicly available on YouTube and were accessed in their original publicly shared form. No personal data were collected. No consent procedures were required, as the study involves analysis of publicly available media texts rather than research with human participants.

## 4. Analysis and Discussion

This chapter analyses three contemporary Pakistani Sufi music videos through a multimodal discourse lens, addressing research questions. The videos examined are Sultan Ateeq Rehman's *Asaa Ik Din Sab Nu Yad Araan* (کلام میاں محمد بخش), Ramzan Jani's *Aj Aakhan Waris Shah Nu* (اج آکھاں وارث شاہ نوں), and Muhammad Samie's *Maikada 3: Bus Ik Nigah* (میکده 3: بس اک نگاہ). Analysis draws on McNeill's (1992) gesture typology and Kress and van Leeuwen's (2021) visual grammar to interpret how bodily and visual modes work together to construct Sufi meaning.

### 4.1 Analysis

#### 4.1.1 Sultan Ateeq Rehman - *Asaa Ik Din Sab Nu Yad Araan*

Sultan Ateeq Rehman's video draws on the classical Punjabi kalam of Mian Muhammad Bakhsh (میاں محمد بخش), a 19th-century Kashmiri Sufi poet whose *Saif-ul-Malook* is among the most revered texts in the Pakistani Sufi canon. The video is shot in outdoor locations, open hilltops, ruined walls, and riverbanks in Azad Kashmir, that code the spatial environment as historically and spiritually charged. Ateeq wears a green turban and long kameez, a sartorial choice that indexes the visual grammar of the classical *darwesh* (Sufi wanderer) and aligns him with a tradition of pious, itinerant sainthood.



*Figure 1. Sultan Ateeq Rehman, Timestamp: 3:20. Upward gaze with closed eyes, encoding hal (spiritual state) during a sustained melodic phrase.*

Throughout the video, Ateeq performs a recurring postural vocabulary centred on the upward gaze. At multiple points during melodically and lyrically intense verses, his head tilts back, his eyes close or look skyward, and his chin rises. This configuration is legible within the Sufi performance convention as an index of *hal* (حال), the transient spiritual state associated with absorption in divine contemplation. Kress and van Leeuwen (2020) observe that gaze direction constructs social relations within an image: an upward gaze positions the subject in relation to a power or ideal located above, encoding aspiration and reverence. Here, it enacts the Sufi theological orientation toward the divine, the practitioner's whole posture becomes a sustained deictic gesture toward God.



**Figure 2.** Sultan Ateeq Rehman, Timestamp: 3:35. Right hand raised, fingers spread and directed outward beside a stone wall - a metaphoric gesture accompanying the phrase referencing the beloved.

Hand gestures in this video fall primarily into two categories identified by McNeill (1992): metaphoric and deictic. When Ateeq sings phrases invoking *yaar* (یار, the Sufi beloved, a multivalent term for both the spiritual guide and God), his right hand rises beside the body with fingers loosely spread, directed outward at mid-chest height. This is a metaphoric gesture that gives spatial form to the concept of the beloved as a present but transcendent entity: the hand neither points precisely at a location nor mimics an action, but enacts the reaching toward something just beyond material grasp. This gesture recurs consistently with linguistically loaded Sufi terms, supporting McNeill's (1992) argument that gesture and speech are co-generated at a single cognitive growth point rather than produced separately.



**Figure 3.** Sultan Ateeq Rehman, Timestamp: 4:30. Both hands raised open, palms forward, supplication gesture during a verse invoking divine mercy.

The most structurally prominent gesture in Ateeq's performance is the bilateral open-palm gesture visible at 4:30 and several other points. Both hands rise to roughly shoulder height with palms forward and open. This is the embodied form of *dua* (دعا, supplication), one of the most conventionalised bodily practices in Islamic devotional life. Its deployment at lexically significant moments, typically when *rab* (رب, Lord) or *rehmat* (رحمت, divine mercy) appear in the *kalam*, creates a semiotic synchrony in which the verbal invocation and the bodily posture of petition reinforce one another. The gesture does not merely illustrate the prayer: within Sufi performance tradition, following Halid (2022), it constitutes the body as a vehicle of sincere appeal, enacting the theological claim that the seeker's whole being is oriented toward the divine.



**Figure 4.** Sultan Ateeq Rehman, Timestamp: 4:50. Upward gaze with one hand open at chest height, encoding spiritual longing and proximity to the divine during a final melodic climax.

The spatial setting amplifies these gestural meanings through visual composition. Kress and van Leeuwen (2020) note that framing choices, what is included in the shot and at what distance, construct the viewer's social relationship to the subject. The camera frequently uses medium close-up shots that include both Ateeq's face and upper body, ensuring that facial expression and hand gesture are simultaneously readable. The open landscape behind him, mountains, sky, ancient stone, codes the performance as part of a spiritually significant geography, embedding the Sufi wanderer's body within a visual field that reinforces the tradition's emphasis on divine immanence in the natural world.

#### 4.1.2 Ramzan Jani - Aj Aakhan Waris Shah Nu

Ramzan Jani's performance of Amrita Pritam's 1947 Punjabi poem is set against the architectural heritage of Sindh and southern Punjab: Mughal-era monuments, shrine gateways, and the open courtyards of dargahs appear throughout. The lyric text oscillates between the vocabulary of Sufi love, *ishq* (عشق), *wanjhli* (وَنجھلی, Ranjha's flute), *kitaab-e-ishq* (کتاب عشق, book of love), and the language of historical grief: *laashaan* (لاشائاں, corpses), *lahu* (لہو, blood). Jani performs standing throughout, wearing a mustard yellow shalwar kameez that contrasts sharply with the terracotta and sand tones of the settings, making his body visually salient within every frame.



**Figure 5.** Ramzan Jani, Timestamp: 0:40. Dargah setting: men in supplication posture around a shrine, establishing the spatial and devotional register of the performance.

The video opens inside a shrine, where attendants stand in collective supplication around the grave of a saint. This establishing shot positions the performance within a specific spatial

semiotics: the dargah (درگاہ) is the site where the living communicates with the spiritually elevated dead, and the poem is itself an invocation addressed to the deceased Sufi saint Waris Shah. The shot encodes the theological premise of wilayat (ولایت), the continuing spiritual authority and presence of the saint, as a spatial reality, the genre's spatial grammar makes the poem's central claim legible before a word is sung.



*Figure 6. Ramzan Jani, Timestamp: 1:30. Standing at a monument with right arm raised and hand extended, deictic gesture on the invocation 'Waris Shah nu'.*

Jani's most distinctive gesture is a deictic arm extension that occurs during each delivery of the refrain's central invocation, Waris Shah nu (وارث شاہ نوں, meaning 'to Waris Shah', calling on the saint). The right arm rises to shoulder level and extends outward with open hand, directed toward the camera and beyond it. McNeill (1992) classifies this as deictic gesture: it points toward a referent, though here the referent is absent (the saint in the grave) and imagined rather than physically present. The gesture creates a spatial axis between the performer and the addressed other, making visible the poem's core dynamic of calling across the boundary between the living and the dead. Kress and van Leeuwen's (2020) concept of the demand image is relevant here: when the hand extends toward the viewer, it recruits the audience into the act of invocation, positioning them as co-participants rather than observers.



*Figure 7. Ramzan Jani, Timestamp: 3:05. Both fists raised at chest height before a shrine door — rhythmic beat gesture during the verse on collective grief (the suffering ones).*

At verses describing collective suffering, *dardmandan* (دردمندان, the suffering ones) and the imagery of flooded rivers and broken communities, Jani's gesture vocabulary shifts. The open-palmed reaching gesture gives way to raised fists held at chest level. McNeill's (1992) beat gesture typology applies here: these rhythmic, bounded movements mark prosodic structure rather than representing specific semantic content. However, in context, the fist shape carries additional weight: tightly held fists at chest height encode emotional containment, grief held in the body rather than released. The shrine door directly behind Jani in this shot adds a visual layer, the gateway to the spiritual world remains closed, framing the grief as petitionary, a plea not yet answered.

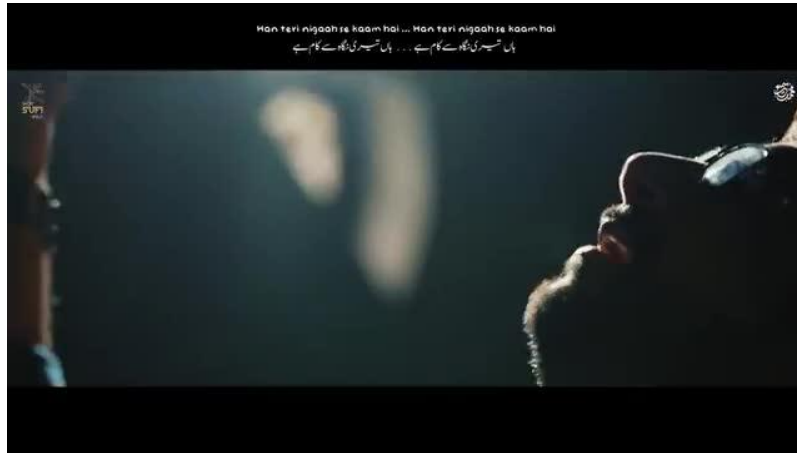


**Figure 8.** Ramzan Jani, Timestamp: 4:15. Open palms at waist level beside a microphone stand, palms-down gesture encoding lamentation on the verse describing corpses in the fields.

When Jani delivers the verse describing the fields of corpses, *bele laashaan vichiyaan* (بیلے لاشاں وچھیاں, fields lined with the dead), both hands open with palms directed downward at waist level. The spatial directionality of this gesture is semiologically precise: where the upward-directed open palm encodes aspiration and petition, the downward palm encodes the earthly, the fallen, the dead. This constitutes a spatial metaphor in McNeill's (1992) terminology, giving physical form to the abstract concept of mass death by locating it below the body's centre. The lyric subtitle visible in the frame reinforces this reading. The contrast between this gesture and the upward invocatory gestures earlier in the video enacts the poem's emotional architecture: a movement between the heights of spiritual aspiration and the depth of historical devastation.

#### 4.1.3 Muhammad Samie - Maikada 3 (Bus Ik Nigah)

Muhammad Samie's Maikada 3 is the most cinematically produced of the three videos, running to approximately fourteen and a half minutes and integrating multiple visual registers: contemporary studio space, warm amber-lit interior settings, black-and-white archival-aesthetic sequences, and a collective mehfil (gathering) scene. Samie describes his work through the label Short Sufi Spells, positioning it as an accessible, condensed delivery of classical Sufi content for a digital audience. The kalam draws on Jigar Muradabadi's classical Urdu verse alongside Samie's own original compositions, and the Maikada (میکدہ, tavern) imagery operates through tawriyah (توریہ, double meaning): the wine-house is the Sufi spiritual gathering, the wine is divine love, and the *saaqi* (ساقی, server of wine) is the spiritual guide.



**Figure 9.** Muhammad Samie, Timestamp: 4:15. Upward profile gaze in low-lit setting with lyric overlay 'Han teri nigaah se kaam hai', the body as a sign of spiritual longing.

Samie's predominant gestural mode is stillness. Unlike the expansive gestures of Ateeq or the rhythmic body engagement of Jani, Samie remains largely immobile for extended portions of the video, allowing facial expression and gaze orientation to carry the primary semiotic load. The profile shot at 4:15 shows his face tilted upward in low light, the lyric overlay reading *Han teri nigaah se kaam hai* (ہاں تیری نگاہ سے کام ہے, 'Yes, it is your glance that I need'). Kress and van Leeuwen (2020) note that an offer image, in which the subject does not directly address the viewer, constructs a contemplative, observational relationship rather than a confrontational one. Here, the upward tilt creates a spatial metaphor of yearning: the seeker's face is oriented toward the divine, which is above and beyond the frame.



**Figure 10.** Muhammad Samie, Timestamp: 11:40. Index finger raised on a phrase about spiritual insight, an iconic Islamic gesture encoding tawhid (divine unity).

The most analytically significant discrete gesture in Samie's performance is the raised index finger visible at 11:40 during a verse about spiritual insight. The extended index finger is among the most semiotically loaded gestures in Islamic cultural practice: it is the physical form of the shahada (شہادہ, declaration of faith) and specifically the affirmation of tawhid (توحید, divine unity). By deploying this gesture at a phrase about clarity of vision and spiritual understanding, Samie aligns the Sufi seeker's experience of divine nearness with the foundational Islamic theological claim of monotheism. This is gesture functioning as what McNeill (1992) calls an

emblematic sign, a conventionalised form carrying culturally specific meaning independent of the speech it accompanies. The lyric subtitle at this moment reinforces the reading: the gesture and the text converge on the same theological claim.



**Figure 11.** Muhammad Samie, Timestamp: 8:20. Collective mehfal scene: two performers in close bodily engagement over shared instruments, encoding the communal dimension of Sufi spiritual seeking.

The collective scene at 8:20 introduces a spatial and relational dimension absent from the solo performance moments. Two performers lean toward one another, bodies overlapping in the frame over a shared instrument. This spatial configuration encodes the relational dimension of Sufi practice: the murshid-murid (master-disciple) relationship in which spiritual transmission occurs through proximity and shared presence. Kress and van Leeuwen (2020) observe that social distance in visual composition, indicated by the amount of the subject's body included in the frame and the spatial relationship between figures, constructs the degree of intimacy or formality of the represented relationship. The tight framing and overlapping bodies construct intimacy, suggesting that the Sufi path is not solitary contemplation but a shared, embodied practice.



**Figure 12.** Muhammad Samie, Timestamp: 9:50. Face close-up during the climactic phrase 'Ke bandagan-e-Khuda par zaban daraz kare', the body at the threshold of wajd.

The close-up of Samie's face at 9:50 shows a configuration associated in Sufi performance convention with the approach of wajd (وجد, spiritual ecstasy): eyes slightly closed, jaw relaxed, the face angled slightly upward and tilted to one side. This is less a discrete gesture than a postural configuration, and its semiotic force comes from its contrast with Samie's usual stillness. The lyric

at this moment — Ke bandagan-e-Khuda par zaban daraz kare (کہ بندگان خدا پر زباں دراز کرے), meaning 'that one should not be presumptuous before the servants of God', is a statement of Sufi adab (spiritual etiquette). The body's visible surrender in this frame enacts the very posture of humility that the lyric prescribes, achieving a semiotic unity between the verbal and corporeal modes that is the hallmark of deeply integrated Sufi performance.

#### 4.1.4 Comparative Gesture Analysis

The following table provides a cross-video comparison of the gesture types identified in each performance, the Sufi concepts they encode, and the verbal anchors with which they are synchronised.

**Table 1: Comparative Gesture Analysis Across Three Pakistani Sufi Music Videos**

Performer	Timestamp	Gesture Type	Description	Verbal Anchor	Sufi Concept Encoded
<b>Sultan Ateeq Rehman</b>	3:20	Postural-meditative	Head tilted back, eyes closed or skyward	Sustained melodic phrase	Hal (حال) - transient spiritual absorption
	3:35	Metaphoric	Right hand raised, fingers spread, directed outward	Yaar (یار, the beloved)	Ishq (عشق) - divine love; longing for the spiritual other
	4:30	Iconic-Emblem	Both palms open and forward at shoulder height	Rab (رب), Rehmat (رحمت)	Dua (دعا) - supplication; faqr (فقر) - spiritual poverty before God
	4:50	Deictic-Metaphoric	Upward gaze, one hand open at chest height	Climactic melodic phrase	Qurb (قرب) - divine nearness; aspiration toward the transcendent
<b>Ramzan Jani</b>	0:40	Spatial-Compositional	Collective supplication at dargah (setting)	Poem's opening invocation	Wilayat (ولایت) - continuing presence of the saint; ziyarat
	1:30	Deictic	Right arm extended toward camera/saint	Waris Shah nu (وارث شاہ نون)	Munajat (مناجات) - petition to a spiritual intercessor
	3:05	Beat / Rhythmic	Fists raised at chest, contained and held	Dardmandan (دردمندان)	Dard (درد) - sacred grief; collective spiritual suffering
<b>Muhammad Samie</b>	4:15	Spatial-Metaphoric	Palms open, directed downward at waist level	Laashaan (لاشان, corpses)	Fana (فنا) - dissolution; the earthly as opposed to the divine
	4:15	Postural-meditative	Profile, face upward in low light, stillness	Nigah (نگاہ, the glance)	Muraqaba (مراقبہ) - deep meditative contemplation
	8:20	Spatial-Relational	Close bodily engagement with co-performer	Collective mehfil setting	Murshid-murid (مرشد-مرید) - master-disciple spiritual transmission
	11:40	Emblematic	Index finger raised, pointing upward	Verse on spiritual insight	Tawhid (توحید) - divine unity; the foundational Islamic declaration
	9:50	Postural-expressive	Head tilted, eyes closing, face relaxed and angled up	Adab verse	Wajd (وجد) - spiritual ecstasy; surrender of the self

#### 4.2 Discussion

#### 4.2.1 The Gesture Vocabulary of Sufi Identity

Taken together, the three videos reveal a shared gestural vocabulary through which Sufi identity is constructed and communicated in contemporary Pakistani music video production. Across all three performances, certain embodied configurations recur with demonstrable consistency: the upward gaze or tilted-back head as an index of spiritual absorption; the open-palm gesture as a sign of supplication and receptivity; and the raised or extended hand as an encoding of address to the divine or to a spiritual intercessor. These recurring forms constitute what might be termed a Sufi gesture grammar, a set of culturally recognised bodily signs whose meanings are stable across different performers, different linguistic registers, and different spatial settings.

This finding responds directly to the first research question. The gestures used in these videos are not arbitrary or idiosyncratic but draw on a shared semiotic reservoir rooted in Islamic devotional practice and Sufi performance tradition. The open palm in dua, the raised index finger of the shahada, the downward gaze of grief and the upward tilt of aspiration: these are forms whose meanings are culturally encoded and immediately legible to audiences socialised within Pakistani Muslim culture. McNeill's (1992) framework helps explain why these forms are so stable: as emblematic gestures carrying conventionalised cultural meanings, they operate semi-independently of speech, maintaining their semantic content across different verbal contexts.

#### 4.2.2 How Gesture Encodes Sufi Concepts

The second research question asks how specific gestures encode key Sufi concepts. The analysis reveals that this encoding operates through three distinct mechanisms. The first is spatial metaphor: the vertical axis of the body is consistently used to organise the Sufi conceptual field, with the divine, the transcendent, and spiritual aspiration located above and the earthly, the deceased, and physical suffering located below. Jani's downward palms on laashaan (لاشآن) and upward extended arm on the saint's invocation are the clearest examples, but Ateeq's upward gaze and Samie's upward profile shot enact the same spatial logic. This is precisely what McNeill (1992) means by metaphoric gesture: the body's spatial orientation gives physical form to abstract theological concepts.

The second mechanism is semiotic synchrony, the co-occurrence of gesture and speech at precisely the moment of a lexically or theologically significant word. The data shows a consistent pattern across all three performers: gesture strokes, in McNeill's (1992) terminology the moment of maximum effort or extent in a gesture, fall on or immediately before key Sufi vocabulary items. Ateeq's bilateral open-palm gesture reaches its fullest extent on rab (رب) and rehmat (رحمت). Jani's arm extends fully on Waris Shah nu (وارث شاه نون). Samie's index finger rises at the moment of the verse's claim about spiritual insight. This synchrony is not accidental: following McNeill's (1992) growth point theory, gesture and speech are produced from a shared cognitive representation, so they peak together at the point of maximum conceptual salience.

The third mechanism is what may be termed performative embodiment, drawing on Halid's (2022) framework: the body does not represent a spiritual state but enacts and transmits it. This is most visible in Samie's stillness and in Ateeq's sustained closed-eye posture. In neither case does the performer illustrate a concept through gesture; instead, the whole body becomes an instance of the concept. The stillness of Samie's muraqaba (مراقبه) is not a sign pointing toward contemplation but an enactment of it. This distinction is theologically significant within Sufi tradition, where the boundary between representing and experiencing a spiritual state is deliberately collapsed.

### 4.2.3 Differences Across Performers and Their Implications

While the three performers share the core gestural vocabulary described above, their deployment differs significantly, and these differences carry interpretive weight. Sultan Ateeq Rehman's gesture vocabulary is the most spatially expansive, with open bilateral gestures and broad arm movements that use the full available space of the outdoor frame. This expansiveness fits the visual grammar of the video, which situates the performer within a vast natural landscape, and constructs a Sufi identity oriented toward the cosmological: the seeker whose body, like the landscape, is scaled to the dimensions of divine creation.

Ramzan Jani's gestures are the most emotionally dynamic, shifting between the reaching-toward of deictic invocation and the downward containment of grief, tracking the poem's oscillation between spiritual aspiration and historical lament. The use of architectural heritage settings, Mughal monuments, shrine gateways, contextualises these gestures within a visual grammar of historical depth and cultural continuity, constructing a Sufi identity that is explicitly rooted in a shared regional and spiritual history. The fact that this is a Pakistani male performer performing a poem by a Hindu Punjabi woman, addressed to an 18th-century Muslim Sufi saint, makes the gestural enactment of cross-border cultural continuity itself a discursively significant act.

Muhammad Samie's performance is the most restrained and the most cinematically mediated. His minimal gesture vocabulary, dominated by stillness and single-finger or profile-face configurations, constructs a Sufi identity that is inward-directed and intellectually self-possessed: the urban, educated seeker whose spiritual life is private and contemplative rather than communally ecstatic. The sophistication of the video production, multi-register visual aesthetics, bilingual lyric subtitles, cinematic lighting, extends this construction, signalling that the tradition is compatible with modern cultural literacy. Kress and van Leeuwen (2020) observe that the modality of a semiotic representation, the degree to which it claims to be real or true, is calibrated through visual production choices: Samie's high-production aesthetic constructs a claim for the tradition's contemporaneity and seriousness.

### 4.2.4 Sufi Identity Through the Body

Across all three videos, the body is the primary site through which Sufi identity is constructed for contemporary Pakistani audiences. This construction operates at three levels simultaneously. At the individual level, each performer's gesture vocabulary constitutes a particular subjectivity within the broader category of Sufi identity: the classical Sufi wanderer (Ateeq), the historically aware devotee (Jani), the contemporary mystic intellectual (Samie). At the traditional level, the shared gestural vocabulary connects all three performances to the established conventions of South Asian Sufi performance, legitimating each video as an authentic continuation of the tradition. At the theological level, the consistent encoding of core Sufi concepts — hal, wajd, dua, muraqaba, tawhid, wilayat — through specific bodily forms translates the abstract theological content of the kalam into immediately legible embodied signs.

This multilevel construction is precisely what Kress and van Leeuwen (2020) mean when they describe multimodal meaning-making as the simultaneous operation of multiple modes, each carrying complementary or extending meanings that together produce a semiotic whole greater than any single mode could achieve alone. The voice, the body, the setting, the camera, and the lyric text are all active participants in the construction of Sufi identity in these videos, and the body's gesture vocabulary is among the most semantically dense of these participants.

#### 4.2.5 Conclusion

This analysis has demonstrated that contemporary Pakistani Sufi music videos deploy a coherent and culturally specific gesture vocabulary through which Sufi identity is constructed and communicated across multiple performers, linguistic registers, and visual contexts. The three videos examined show that gestures are not supplementary to the spiritual meanings of Sufi kalam but are among their primary semiotic vehicles. The upward gaze, the open-palm supplication, the deictic arm extension, the emblematic raised index finger, and the postural configurations of hal and wajd all encode specific Sufi theological concepts with a consistency that reflects the deep cultural embeddedness of this gesture grammar in Pakistani devotional life.

The comparison across performers also reveals that within this shared vocabulary, significant individual variation constructs distinct Sufi subject positions oriented toward different aesthetic and theological emphases. The classical spatial expansiveness of Sultan Ateeq Rehman, the emotionally dynamic and historically rooted gestures of Ramzan Jani, and the restrained, contemplative stillness of Muhammad Samie together demonstrate that Sufi identity in contemporary Pakistani music video is not monolithic but a field of positions united by common semiotic resources and differentiated by their specific orchestration.

These findings have implications beyond this study. They suggest that gesture analysis offers a productive method for examining how spiritual traditions are maintained and adapted through digital media, and that the Sufi music video genre warrants sustained multimodal attention as a significant site of cultural and theological meaning-making in contemporary Pakistan. Future research might extend this framework to female Sufi performers, to non-Pakistani South Asian Sufi video production, and to audience reception studies examining how the gesture vocabulary of these videos is interpreted by viewers across different cultural contexts.

### 5. CONCLUSION AND FINDINGS

#### 5.1 Findings

The findings of this study are presented below in direct response to each of the two research questions that guided the investigation.

##### 1. *What gestures and body language are used in contemporary Pakistani Sufi music videos?*

The analysis identified a consistent and culturally recognisable gestural vocabulary across all three videos. This includes: the upward-tilted head with closed or skyward gaze encoding hal (transient spiritual absorption); the bilateral open-palm gesture at shoulder height encoding dua (supplication) and faqr (spiritual poverty before God); the extended deictic arm directed toward an imagined spiritual presence encoding munajat (petition to a saint or the divine); the downward open palm encoding grief and the earthly; the raised index finger as an emblematic sign of tawhid (divine unity); and postural configurations of sustained stillness encoding muraqaba (meditative contemplation). These gestures draw on both Islamic devotional conventions and established Sufi performance tradition, rendering them immediately legible to audiences socialised within Pakistani Muslim culture. Individual performers also exhibit idiosyncratic bodily elements, Ateeq's spatially expansive outdoor movements, Jani's rhythmic fist gestures encoding collective grief, and Samie's restrained facial expressivity, yet all operate within the boundaries of a shared gestural grammar.

## ***2. How do these gestures and body language encode key Sufi concepts across these music videos?***

Gesture encodes Sufi theological concepts through three analytically distinct mechanisms. First, spatial metaphor: the vertical axis of the body consistently positions the divine, the transcendent, and spiritual aspiration above and the earthly, the deceased, and physical suffering below, giving physical form to abstract theological distinctions. Second, semiotic synchrony: gesture strokes peak at lexically and theologically significant Sufi vocabulary items, *rab*, *rehmat*, *Waris Shah nu*, confirming McNeill's (1992) growth point theory and demonstrating that gesture and speech are co-generated from a shared cognitive representation. Third, performative embodiment: rather than merely representing spiritual states, gestures such as Samie's stillness during *muraqaba* and Ateeq's sustained upward gaze during *hal* enact those states rather than signifying them. Together, these mechanisms translate the abstract theological vocabulary of classical *kalam* into visible, embodied signs accessible to contemporary digital audiences.

### **5.2 Conclusion**

This study has demonstrated that gesture and body language are not peripheral to Sufi music video performance but are among its primary vehicles of theological meaning. Across three contemporary Pakistani Sufi music videos, a coherent gestural vocabulary encodes core Sufi concepts, *hal*, *wajd*, *dua*, *tawhid*, *wilayat*, *muraqaba*, with a consistency rooted in centuries of devotional practice. Performer-specific variation simultaneously constructs distinct Sufi subjectivities: the classical *darwesh* (Ateeq), the historically rooted devotee (Jani), and the contemplative contemporary mystic (Samie). The body, situated within carefully composed visual frames, operates as a primary semiotic mode through which a living spiritual tradition is sustained, adapted, and communicated in the age of algorithmically mediated content. These findings extend the theoretical claims of both social semiotics and gesture theory into the underexplored domain of non-Western religious media, demonstrating that the Sufi body carries theological weight that no single analytical mode can adequately capture in isolation.

### **5.3 Recommendations**

Future research should apply this framework to female Sufi performers in Pakistan, where gendered performance conventions may produce distinct gestural vocabularies and different constructions of spiritual authority. The framework should be extended to Sufi music video production in India and Bangladesh to determine whether the gestural grammar identified here is specific to Pakistani performance or broadly characteristic of South Asian Sufi media. Empirical reception studies examining how audiences across different age groups, cultural backgrounds, and levels of religious literacy interpret the gestural vocabulary identified here would productively complement the production-focused analysis of this study. Sufi music video producers should recognise that gesture visibility and camera framing carry direct semantic weight; conscious attention to the synchrony between gestural and verbal modes strengthens the theological coherence of the final production.

### **5.4 Implications**

This study offers a replicable model for applying gesture typology and visual grammar to mediated religious performance in non-Western contexts, extending these frameworks beyond the Western conversational and print settings in which they were originally developed. Sufi music videos are significant sites of cultural and theological meaning-making whose gestural dimension has been substantially overlooked; the genre warrants sustained multimodal scholarly attention



within Pakistani media research. Gesture is a critical mechanism through which spiritual traditions sustain coherence in digital formats, with broad implications for how religious communities transmit devotional practices and communal identities across generations and geographies.

## References

- Bunt, G. R. (2018). *Hashtag Islam: How cyber-Islamic environments are transforming religious authority*. University of North Carolina Press.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- Denzin, N. K., & Lincoln, Y. S. (2018). *The SAGE handbook of qualitative research* (5th ed.). SAGE Publications.
- Fairclough, N. (2013). *Critical discourse analysis: The critical study of language* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315834368>
- Frembgen, J. W., & Ripken, J. (2011). At the shrine of the red Sufi: Five days and nights on pilgrimage in Pakistan.
- Halid, R. I. B. R. (2022). Sufism, spiritual performativity and theological contestations: Parallelisms of the Aceh and Perak sultanates across the centuries. *International Journal of Asia-Pacific Studies*, 18(2), 309–338. <https://doi.org/10.21315/ijaps2022.18.2.13>
- Jabbar, Q., Gul, A., & Ashraf, J. (2025). Power of visuals in political branding: A multimodal discourse analysis of visualatory campaigns. *Wah Academia Journal of Social Sciences*, 4(1), 1516-1528.
- Jani, R. (2024, January 15). Aj aakhan Waris Shah nu | Sufi kalam 2024 | Amrita Pritam [Video]. YouTube. <https://www.youtube.com/watch?v=dXTifGkgBVQ>
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.
- Kress, G., & Van Leeuwen, T. (2020). *Reading images: The grammar of visual design*. Routledge.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.
- Mir, F. (2010). *The social space of language: vernacular culture in British colonial Punjab* (Vol. 2). Univ of California Press.
- Qayyum, N., & Ali, M. J. (2024). Notion of beauty in the selected Pakistani advertisements: A multimodal critical discourse analysis. *VFAST Transactions on Education and Social Sciences*, 12(1), 37-51.
- Qureshi, R. (1986). *Sufi music of India and Pakistan: sound, context and meaning in Qawwali* (Vol. 1). CUP Archive.
- Rehman, S. A. (2023). Asaa ik din sab nu yad araan | New Kalma Sharif | Kalam Mian Muhammad Bakhsh Part 2 [Video]. YouTube. <https://www.youtube.com/watch?v=o07YSAPa1Aw>
- Sakata, H. L. (1994). The Sacred and the Profane: "Qawwālī" Represented in the Performances of Nusrat Fateh Ali Khan. *The World of Music*, 86-99.
- Samie, M. (2024, August 9). Maikada 3 (Bus ik nigah) | Official video [Video]. YouTube. <https://www.youtube.com/watch?v=FXOOeSA10fw>
- Sarrazin, N. (2013). Devotion Or Pleasure?: Music and Meaning in the Celluloid Performances of Qawwali in South Asia and the Diaspora. In *Music, Culture and Identity in the Muslim World* (pp. 178-199). Routledge.
- Taj, I., & Ameen, M. (2024). CULTURAL REFLECTIONS IN PAKISTANI RAP MUSIC: A STUDY OF 'THE SIBBI SONG' AND 'KARACHI CHAL'. *Journal of Social Sciences and Humanities*, 63(1), 23-58.
- Van Leeuwen, T. (2023). Social semiotics and multimodality. In L. Pauwels & D. Mannay (Eds.), *The SAGE handbook of visual research methods* (2nd ed., pp. 441–451). SAGE Publications. <https://doi.org/10.4324/9781003082637-28>
- Werbner, P. (2003). *Pilgrims of love: The anthropology of a global Sufi cult*. Hurst & Company.
- Williams, R. D., & Mahmood, R. (2019). A soundtrack for reimagining Pakistan? Coke Studio, memory and the music video. *BioScope: South Asian Screen Studies*, 10(2), 179–203. <https://doi.org/10.1177/0974927619896771>