TOFEDU: The Future of Education Journal
Volume 5 Number 4 (2026) Page: 6220-6226
E-ISSN 2961-7553 P-ISSN 2963-8135
https://journal.tofedu.or.id/index.php/journal/index

Multimodal Discourse Analysis of American Sign Language Representation in Sia's "Soon We'll Be Found" Music Video

Sabrina Zaliana Putri
sabrinaaazp@gmail.com
English Literature Department, Faculty of Adab and Humanities,
Universitas Islam Negeri Sulthan Thaha Saifuddin Jambi, Jambi, Indonesia

ABSTRACT
This study examines how American Sign Language (ASL) is represented and functions as a semiotic mode in Sia's "Soon We'll Be Found" music video. It aims to identify how sign language interacts with other multimodal elements and to explain the meanings, emotions, and visual messages conveyed through this interaction. The study employs a descriptive qualitative method and applies Kress and van Leeuwen's (2001) four strata framework of Multimodal Discourse Analysis (MDA): discourse, design, production, and distribution. The data consist of selected visual frames showing sign language performance, song lyrics, audiovisual elements, and public responses on YouTube. Data were collected through repeated observation, timestamping, frame capture, documentation, and contextual verification of signs using sign-language references. The findings indicate that ASL functions as a primary semiotic mode rather than a decorative visual element. Its interaction with lyrics, music, facial expressions, body movements, gaze, camera framing, lighting, and color constructs a coherent visual narrative. The signs communicate ideas about communication, emotional connection, the desire to be understood, and interpersonal relationships, while expressing loneliness, vulnerability, uncertainty, hope, and the desire for connection. The study also finds that YouTube reception extends the video's message of inclusivity to Deaf and hearing audiences. It concludes that sign language in this music video operates as a substantive communicative resource that connects artistic expression with social inclusivity.

Keywords: American Sign Language; inclusivity; music video; multimodal discourse analysis; Sia

INTRODUCTION

Language is a fundamental resource for human communication and the construction of social meaning. Communication, however, is not limited to spoken or written words. Meaning can also be conveyed through visual and bodily resources, including gestures, facial expressions, movement, and spatial arrangements. Sign language is an important example of this visual-gestural form of communication. It is a structured language used by Deaf communities and has its own linguistic resources rather than functioning merely as a collection of gestures or a visual substitute for spoken language (Stokoe, 2005; Supalla & Cripps, 2017).

In audiovisual media, sign language can operate together with images, music, lyrics, movement, and cinematography. These modes contribute different resources to the construction of meaning. Kress and van Leeuwen (2001) explain that communication is multimodal because meaning is produced through the interaction of multiple modes. From this perspective, a music video should not be understood only through its lyrics or soundtrack. Its visual and gestural elements also participate in communicating the message. Sign language therefore deserves attention not only as a linguistic system but also as a semiotic resource within popular media.

This work is licensed under a Creative Commons Attribution 4.0 International Licens

E-ISSN 2961-7553
P-ISSN 2963-8135

The inclusion of sign language in popular music has become a visible way of addressing accessibility and representation. Sari (2022), for example, examined the representation of sign language in BTS's "Permission to Dance" and showed how sign language can communicate visual messages in popular music. Research on multimodal communication has also emphasized the importance of framing and non-manual markers in the presentation of sign language (Aditya & Pradotokusumo, 2022). More broadly, multimodal studies of music videos demonstrate that lyrics, sound, gestures, and cinematography can work together to create meanings that are not fully conveyed by any single mode (Oyebamiji & Olutayo, 2025).

Nevertheless, the presence of sign language in a music video does not, by itself, explain how it functions in the overall construction of meaning. When sign language is treated only as a visual addition, the relationship between the signs, facial expressions, body movements, lyrics, musical rhythm, and cinematic choices may be overlooked. Questions of representation and inclusivity also require attention to how sign language is positioned within the visual narrative and how audiences encounter it. Thompson (2021) raises a related concern about the aestheticization of disability in popular culture, highlighting the importance of examining whether representation has a substantive communicative role rather than serving only as a visual effect.

Sia's "Soon We'll Be Found" provides an opportunity to examine these issues through a single music video. The video incorporates American Sign Language (ASL) within a melancholic and introspective audiovisual setting. Its signing performances appear alongside facial expressions, body movements, lyrics, music, lighting, color, and camera framing. The interaction of these elements makes the video a multimodal text in which sign language contributes to both the communication of lyrics and the expression of emotional and social meanings.

Previous research has examined sign language in popular media, multimodal framing, and the interaction of semiotic modes. However, the present study focuses specifically on the representation of ASL in Sia's "Soon We'll Be Found" through the four strata of Multimodal Discourse Analysis. It considers not only the meanings of selected signs but also how their visual and auditory surroundings contribute to the video's narrative and how the video circulates among audiences on YouTube. This focus addresses the need for a more integrated examination of sign language as a central semiotic mode rather than as a decorative component.

Accordingly, this study addresses two research questions: (1) How is sign language used through multimodal discourse analysis in Sia's "Soon We'll Be Found" music video? and (2) What meanings, emotions, and visual messages are conveyed through the use of sign language in the music video? The objective is to analyze the multimodal elements involved in the representation of ASL and to explain its contribution to the construction of meaning, emotion, and visual messages. The study's contribution lies in applying the four-strata MDA framework to examine sign language across discourse, design, production, and distribution in one music video.

RESEARCH METHOD

This study uses a descriptive qualitative research design. This approach was selected because the research aims to interpret the meanings and communicative functions of sign language in an audiovisual text rather than to calculate statistical relationships or make numerical generalizations. The researcher acts as the primary instrument in observing, documenting, and interpreting the semiotic resources in the music video. The analysis is guided by Kress and van Leeuwen's (2001) Multimodal Discourse Analysis framework, which consists of four interconnected strata: discourse, design, production, and distribution.

This work is licensed under a Creative Commons Attribution 4.0 International Licens

E-ISSN 2961-7553
P-ISSN 2963-8135

The primary research object is Sia's "Soon We'll Be Found" music video. The analysis uses the official VEVO version released on October 5, 2009, with a duration of approximately four minutes, as identified in the thesis. The data include visual scenes in which ASL is performed, the corresponding song lyrics and audio, and the visual resources accompanying the performance. These resources include hand movements, facial expressions, body posture, gaze direction, camera framing, lighting, color, and spatial composition. Public responses in the YouTube comments section are also considered in examining the distribution stratum.

Data were collected through repeated audiovisual observation and visual documentation. First, the researcher watched the video repeatedly to understand its overall narrative and atmosphere. Second, the timestamps of signing performances were recorded to distinguish signs with a communicative function from movements used only as choreography. Third, key frames were captured at high resolution to document hand configurations and non-manual markers. Fourth, the researcher identified accompanying elements, including lyrics, musical rhythm, background color, lighting, and camera focus. The visual records and observations were then organized chronologically for analysis. Finally, the observed hand movements were cross-referenced with sign-language references to verify their lexical meanings and contextual use.

Data analysis followed the four strata proposed by Kress and van Leeuwen (2001). The discourse stratum examined the broader social and communicative ideas represented in the video, including communication, emotional isolation, connection, and inclusivity. The design stratum examined how signs were selected and coordinated with lyrics, music, and visual composition. The production stratum focused on the physical realization of the signs, including hand movement, facial expressions, body movement, lighting, and camera techniques. The distribution stratum considered how the finished video was circulated through YouTube and how public comments reflected audience responses. The analysis also considered intersemiotic complementarity, or the way different modes reinforce and extend one another (Royce, 1998).

RESULTS AND DISCUSSION

Representation of ASL through Multimodal Elements

The findings show that ASL functions as a primary semiotic mode in "Soon We'll Be Found," rather than appearing merely as a decorative visual element. The signing is integrated into the progression of the song and is presented through the interaction of hand gestures, facial expressions, body movements, gaze direction, shot framing, color, lyrics, and music. The use of close-up and medium shots makes hand movements and facial expressions visible, while the positioning of the performer within the frame gives the signing space a clear visual role. These elements work together to make sign language an important part of the video's visual narrative.

The representation of ASL is connected to the structure of the song, including its verses, chorus, and outro. The signs do not appear as isolated movements; instead, they are coordinated with the lyrics and the musical progression. This relationship allows the gestural mode to reinforce the verbal and auditory modes. The signs communicate key ideas while the surrounding visual and audio elements provide additional contextual and emotional information. Consequently, the video constructs meaning through the combined operation of several semiotic resources.

The Four Strata of Multimodal Discourse

Discourse Stratum

At the discourse level, the video represents emotional isolation and difficulties in interpersonal communication, while also expressing a desire for understanding and human

This work is licensed under a Creative Commons Attribution 4.0 International Licens

E-ISSN 2961-7553
P-ISSN 2963-8135

connection. Sign language is positioned as a meaningful communicative resource through which these themes become visible. The use of ASL also contributes to a discourse of inclusivity by making the song's message accessible through a visual language and by recognizing Deaf audiences as part of the potential audience for popular music.

The signs identified in the analysis include COME ALONG, DAY, SLEEP, LIE, WE, and HAPPY. Their meanings contribute to the development of the video's narrative. For example, COME ALONG and DAY help establish the lyrical and situational context, while WE and HAPPY contribute to meanings of collective solidarity and emotional restoration. The discourse is therefore not limited to literal translation of the lyrics; it also involves broader ideas about vulnerability, relationships, and the possibility of connection.

Design Stratum

At the design level, the video coordinates the gestural mode with lyrics, musical rhythm, and visual composition. The timing of the signs in relation to the lyrics creates intersemiotic complementarity: the visual signs and verbal content reinforce one another, while music and visual atmosphere contribute further layers of meaning. The melancholic piano accompaniment and dim visual scheme support the introspective tone of the song. In this way, the design organizes different modes into a coherent aesthetic structure rather than presenting them as unrelated components.

The signing space also contributes to the construction of meaning. The contrast between a centralized spotlight and a dark, minimalist background creates a visual metaphor of movement from isolation toward human connection. The spatial arrangement draws attention to the performer and to the signs being produced. This design choice strengthens the role of ASL as a central element of the visual narrative.

Production Stratum

At the production level, meaning is realized through the physical performance of signs and the technical choices used to record them. Handshape, location, movement, and orientation provide the manual components of ASL, while facial expressions and gaze function as non-manual markers that contribute to the interpretation of the signs. The performer's facial expressions and body movements intensify the emotional qualities of the song, including loneliness, vulnerability, uncertainty, and hope.

Camera framing, lighting, and color also influence how the signs are perceived. Close-up and medium shots make hand movements and facial expressions more observable. The video's lighting design, including the use of ultraviolet and neon visual effects described in the thesis, draws attention to the hand configurations and signing performance. These production choices help transform the conceptual design into a visible audiovisual performance.

Distribution Stratum

At the distribution level, YouTube functions not only as a platform for sharing the music video but also as a space in which audiences respond to and reinterpret its message. The thesis examines public reception through a purposive sample of 20 viewer comments. The discussion identifies two broad patterns in these responses. First, comments from Deaf viewers or viewers engaging with Deaf culture recognize the importance of the signing and its perceived accuracy. Second, hearing audiences express appreciation and develop interest in sign language after encountering it in the video.

These responses indicate that the circulation of the music video can extend the message of inclusivity beyond the original audiovisual text. The presence of sign language invites audiences to recognize a visual language and to consider Deaf people as participants in popular cultural expression. However, the comments represent online responses and should not be treated as a substitute for direct interviews or focus-group research with Deaf audiences.

This work is licensed under a Creative Commons Attribution 4.0 International Licens

E-ISSN 2961-7553
P-ISSN 2963-8135

Meanings, Emotions, and Visual Messages

The interaction of sign language with other modes conveys ideas about communication, emotional connection, the desire to be understood, and interpersonal bonds. The hand gestures communicate lexical meanings, while facial expressions, body posture, gaze, music, and visual composition give these meanings emotional depth. The emotions identified in the video include loneliness, vulnerability, uncertainty, hope, and the desire for connection. Sign language therefore does more than reproduce the lyrics visually: it gives visible form to feelings and relationships embedded in the song.

The analysis also identifies artistic modifications in some performances, including poetic adaptations of handshape, rhythmic repetition adjusted to the music, and spatial compression to fit the camera frame. These variations were not included in the primary analytical corpus, which focused on standard ASL representations. Their presence nevertheless suggests an area for further research into artistic sign language, poetic license, and performance variation. The distinction is important because the present study focuses on sign language as a semiotic resource in the video, not on providing a comprehensive linguistic analysis of ASL phonology, morphology, or syntax.

Discussion

Taken together, the findings demonstrate that the meaning of "Soon We'll Be Found" is constructed through the coordinated use of gestural, textual, auditory, and visual modes. This supports the central premise of MDA that meaning is not produced by one mode in isolation (Kress & van Leeuwen, 2001). ASL provides a visible communicative structure, while lyrics, music, facial expressions, body movements, framing, and lighting reinforce and extend its meanings. The video can therefore be understood as a multimodal text in which sign language actively shapes the visual narrative.

The findings also show that the signs retain identifiable lexical meanings while participating in an artistic performance. Signs such as DAY, SLEEP, LIE, COME ALONG, WE, and HAPPY communicate recognizable content, but their placement within the song and their performance through facial expression and spatial composition produce additional emotional and social meanings. In particular, WE and HAPPY are interpreted as expressing solidarity, vulnerability, and emotional restoration. This illustrates how a sign's communicative contribution can be expanded by the modes that accompany it.

Finally, the distribution analysis suggests that digital circulation can support the visibility of sign language and encourage audience engagement with inclusivity. The YouTube comments examined in the thesis point to recognition of signing and increased interest in sign language among hearing viewers. These findings should be understood within the limits of the selected video and comment sample. They do not establish the experiences or views of the entire Deaf community, but they show how a popular music video can create a public space for engagement with sign language.

CONCLUSION

Conclusion

This study concludes that American Sign Language in Sia's "Soon We'll Be Found" functions as a primary semiotic mode that contributes to the construction of the music video's visual narrative. Through the four strata of Multimodal Discourse Analysis discourse, design, production, and distribution the study demonstrates that sign language interacts with lyrics, music, facial expressions, body movements, gaze, camera framing, lighting, and color to create a coherent multimodal text. The signs communicate ideas about communication, emotional connection, the desire to be understood, and interpersonal relationships, while

This work is licensed under a Creative Commons Attribution 4.0 International Licens

E-ISSN 2961-7553
P-ISSN 2963-8135

contributing to the expression of loneliness, vulnerability, uncertainty, hope, and the desire for connection.

The analysis further indicates that the video's visual design and production reinforce the communicative function of ASL. The coordination of signs with lyrics and music, the use of a dark and minimalist background, and the emphasis created by camera framing and lighting all contribute to the visibility and emotional force of the signing performance. At the distribution level, the examined YouTube comments suggest that the video supports recognition of sign language and encourages interest among hearing audiences. Overall, the integration of ASL connects artistic expression with a message of social inclusivity.

Limitations and Recommendations

This study is limited to one music video and to an analysis of its audiovisual text and selected YouTube comments. It does not include direct interviews or focus group discussions with Deaf participants, and it does not undertake a full linguistic analysis of ASL structures. Therefore, its findings should not be generalized to all music videos or to the experiences of the wider Deaf community.

Future research may involve direct participation from Deaf audiences through interviews or focus-group discussions to examine the reception and sociocultural impact of sign-language-based music videos. Further studies may also compare music videos across musical genres or investigate artistic modifications of ASL through frameworks concerned with artistic sign language, poetic license, and sociolinguistic variation. For creative industry practitioners, the findings support integrating sign language into the main visual design from the pre-production stage and involving qualified sign-language professionals in the process.

ACKNOWLEDGMENT

The author would like to express sincere gratitude to the supervisors, lecturers, and staff of the Faculty of Adab and Humanities, Universitas Islam Negeri Sulthan Thaha Saifuddin Jambi, for their guidance and support throughout the research process. The author also thanks her family and friends for their encouragement. (Please confirm whether the journal requires this section and whether any additional funder should be acknowledged.).

REFERENCES

Aditya, D., & Pradotokusumo, A. (2022). Multimodal analysis of sign language in digital content: Framing and non-manual markers. Journal of Visual Literacy, 41(2), 88–105.

Baldry, A., & Thibault, P. J. (2006). Multimodal transcription and text analysis: A multimedia toolkit and coursebook. Equinox.

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE Publications.

Jewitt, C. (Ed.). (2014). The Routledge handbook of multimodal analysis (2nd ed.). Routledge.

Kress, G. (2010). Multimodality: A social semiotic approach to contemporary communication. Routledge.

Kress, G., & van Leeuwen, T. (2001). Multimodal discourse: The modes and media of contemporary communication. Arnold.

O'Halloran, K. L. (2011). Multimodal discourse analysis. In K. Hyland & B. Paltridge (Eds.), The Bloomsbury companion to discourse analysis (pp. 120–137). Bloomsbury.

Oyebamiji, G. A., & Olutayo, G. O. (2025). A multimodal discourse analysis of the music video "IBA". Multimodal Communication, 14(2), Article 100582.

Royce, T. D. (1998). Synergy on the page: Exploring intersemiotic complementarity in page-based multimodal texts. JASFL Occasional Papers, 1(1), 25–49.

This work is licensed under a Creative Commons Attribution 4.0 International Licens

E-ISSN 2961-7553
P-ISSN 2963-8135

Sari, A. L. (2022). Representasi makna sign language dalam music video BTS 'Permission to Dance' (Analisis semiotika Charles Sanders Peirce). Jurnal Spektrum Komunikasi, 9(2), 154–168.

Stokoe, W. C., Jr. (2005). Sign language structure: An outline of the visual communication systems of the American deaf. The Journal of Deaf Studies and Deaf Education, 10(1), 3–37.

Supalla, S. J., & Cripps, J. S. (2017). Linguistic accessibility and deaf children. Oxford University Press.

Thompson, R. (2021). The aestheticization of disability in popular culture: Inclusivity or commodification? Disability & Society, 36(8), 1245–1263.

This work is licensed under a Creative Commons Attribution 4.0 International Licens

E-ISSN 2961-7553
P-ISSN 2963-8135