

Mechanisms of Viral Dissemination of Regional Languages Songs on Short-Video Platforms

Manyi Li ^{1*}

1 Changxin International College of Art, Yunnan University, Kunming, Yunnan, P. R. China, 1669275880@qq.com

*Correspondence: 1669275880@qq.com

Abstract

Short-video platforms have fundamentally reshaped the landscape of music dissemination. Taking TIKTOK as a case study, this research examines four regional languages songs that have achieved widespread popularity, aiming to uncover the underlying mechanisms driving their rapid spread. The findings indicate that the dissemination of these songs does not depend on listeners' comprehension of the lyrics. Rather, they share a common dissemination logic: firstly, platform algorithms offer these songs an initial chance to be discovered; next, iconic musical snippets then serve as sonic hooks that make them instantly memorable; then easily imitated visual movements become participatory templates that invite user engagement; and, ultimately, these elements work together to evoke emotional resonance that transcends linguistic barriers.

Keywords: Regional languages songs; Short-video platforms; Music dissemination; Algorithmic recommendation; User participation



1. Introduction

1.1 Research Background and Problem Statement

The emergence of short-video platforms has fundamentally reshaped how music is disseminated world-wide. These platforms are no longer mere entertainment outlets but have become leading channels for music promotion. On Douyin, hundreds of millions of users scroll through videos daily and are constantly exposed to musical content, whether through passive background listening or active engagement with music-driven content. A single song can garner hundreds of millions of plays overnight through a 15-second clip—a scale of dissemination unimaginable in the era of traditional media[1].

Prior research has extensively examined music dissemination on Douyin and similar platforms, identifying key drivers such as immersive viewing experiences, algorithmic recommendation, user imitation, and social sharing.[2] These forces have accelerated and broadened the reach of music, turning it from a purely auditory experience into an audio-visual phenomenon and granting unprecedented exposure to original compositions, obscure old songs, and traditional folk music. [3] At the same time, scholars have documented a range of accompanying problems: inadequate copyright protection, the clip-based fragmentation of songs, stylistic homogenization, a tendency to prioritize entertainment over artistic depth, the fleeting popularity of viral hits, and a growing dissociation between widely known songs and their unrecognized creators.[4] Existing studies have largely addressed these issues from the angles of dissemination characteristics, technological affordances, aesthetic influence, and copyright governance. While these concerns remain significant, they do not directly account for the mechanism through which regional languages songs manage to circulate widely, the interplay of audio-visual participation, algorithmic amplification, and emotional bonding in this specific context warrants a focused investigation.[5]

However, the dissemination of regional languages songs remains an under-explored area. Even when listeners cannot understand the lyrics, these songs can still spread widely. This observation raises several questions: why these songs stand out amid the vast amount of short-video content, whether and how the music generates an emotional appeal that transcends language barriers, and what role platform recommendation algorithms play in driving this phenomenon. To address these questions, this study conducts a qualitative analysis of four widely circulated regional languages songs on Douyin, aiming to uncover the shared logic behind their dissemination. In this study, “less commonly spoken languages” is operationally defined as languages that are neither natively spoken nor widely learned among Douyin’s primary user base in China.

1.2 Analytical Framework, Case Selection, and Method

This study examines the rapid dissemination of these songs through four interrelated dimensions: the driving force of platform algorithms.[6] The acoustic appeal of the sound clips, the participatory nature of the visual content.[7] And the emotional resonance that transcends linguistic barriers.[8] Specifically, acoustic appeal focuses on rhythm, melodic fragments, and timbre as immediately graspable sonic properties, while emotional resonance centers on listeners’ subjective affective responses and their discursive expressions in comments and re-creations. [9]

Four representative songs were selected based on three criteria: (1) the song is in a language unfamiliar to the vast majority of Douyin users in China; (2) it has accumulated over 200 million cumulative views under its dedicated hashtag on the platform; and (3) its wide circulation is demonstrably tied to a distinctive dance challenge or visual imitation trend. The selected cases are as follows. The Vietnamese song “See Tinh”, performed by Hoàng Thùy Linh and produced by the DTAP team, was released on February 20, 2022, and gained widespread circulation on Douyin accompanied by the highly recognizable “Ear-Covering Dance”. The Russian song “Зая” (also transliterated as “Zaya”) by the group Bambinton, released on February 17, 2017, later swept short-video platforms along with the associated “Zaya Dance”. The Thai song “Whatev-

er” by the girl group PIXXIE, released on September 28, 2022, attracted massive attention in 2024 through its high-energy chorus and the accompanying hand-gesture challenge. The Spanish song “Montagem Nada Tropica” by producer Eternxlkz, released on April 18, 2025, circulated widely via a sped-up version and the accompanying “Sachi Shake” dance.

Spanning a release period from 2017 to 2025, these four pieces are in languages largely unfamiliar to most Chinese Douyin users. Despite their divergent linguistic origins and musical styles, all have accumulated hundreds of millions of views and display remarkably similar patterns of widespread circulation, making them well-suited for the comparative analysis undertaken in this study.

To analyze these cases, the study draws on data collected directly from Douyin. For each of four songs, the author identified the primary dedicated hash tag (e.g., “SeeTinh”) and, as of 1 March 2026, retrieved the top 50 user-generated videos ranked by like count under that tag. The associated comment sections were then sampled from the five most-liked videos within each group of 50. A thematic coding approach was applied to the collected sonic, visual, and textual data to identify recurring motifs. Specifically, this involved open coding of user comments to capture recurring expressions of emotional response (e.g., “I feel like crying”, “so healing”) and participatory intention (e.g., “I want to learn this dance”, “here’s my version”), followed by an iterative grouping of these codes into broader themes. By examining how sonic features, dance movements, and user engagement data interact, the analysis uncovers the shared dissemination patterns driving the viral spread of these four songs[10].

2. Douyin’s Dissemination Ecosystem

2.1 Decentralized Algorithmic Mechanism

Content visibility on Douyin is largely governed by a tiered recommendation algorithm.[11] Initially, an uploaded video is shown to a small, randomly assigned seed pool of users. Its subsequent fate is determined not by the creator’s prior fame, but by quantitative engagement metrics generated in this pool—including completion rate, likes, comments, and shares. Videos that surpass certain algorithmic thresholds are then circulated to larger user pools, creating a cascading exposure effect. This feedback-driven, rather than editor-centric, process is commonly referred to as a decentralized recommendation mechanism.[12]

For this study, this logic constitutes the technological precondition for the widespread dissemination of regional languages songs. Since their appeal is embedded not in lyrical meaning but in sonic and visual elements, these songs can accumulate positive engagement signals in seed pools irrespective of listeners’ linguistic backgrounds. In this way, the algorithm systematically lowers the barriers of language and national borders, offering these otherwise niche musical pieces a viable path to mass discovery.

2.2 Platform Features that Promote Participation

Douyin’s interface is designed to lower the barriers to imitation and derivative creation, offering several built-in functions that turn users from passive listeners into active disseminators.

The editing tool allows users to extract the most infectious snippet of a song—often a brief melodic or rhythmic hook, rather than a lyrical passage—and pair it with their own video.[13] This separates the song’s immediate sonic appeal from its verbal meaning, which is especially important for regional languages songs: the snippet can circulate as pure sound, without listeners needing to understand the words. Moreover, recording and template-based reuse functions further turn listening into visible participation.[14] Users can film themselves alongside an existing video or creatively adapt its visual elements, inserting themselves directly into the song’s ongoing spread.

Most importantly, topic-based challenges organize this participatory activity on a large scale.[15] A specific dance move or creative format is designated as the template, which significantly reduces the effort needed to come up with an original idea. Users can feel a sense of participation and creativity simply by



replicating and personalizing the given format, while the hashtag mechanism gathers all these replicas into a single, constantly growing space of circulation.

Together, these features transform a song into a compact, easily replicable combination of sound and movement. They lower the creative threshold while keeping the content highly imitable, which is precisely what allows regional languages songs to spread quickly across language barriers.

3. Case Analysis

3.1 “See Tình” (Vietnam)

Released by Vietnamese singer Hoàng Thùy Linh in 2022, “See Tình” blends electronic dance music with traditional Southeast Asian elements. A closer examination of its musical structure reveals why it was peculiarly suited to rapid, large-scale dissemination on short-video platforms. [16]

The track is built on a 4/4 time signature at 140 BPM, driven by a steady four-on-the-floor kick pattern that, together with additional percussive accents, creates a sense of relentless, evenly distributed pulse. This forward-driving energy maps easily onto visual editing: no matter where a clip is cut, the resulting movement tends to land on a strong beat, substantially lowering the barrier to music-choreography synchronization for participatory content.

The chorus hook—the playful, onomatopoeic “ding-ding-dong-dong”—functions as a pure sonic trigger. For non-Vietnamese listeners, any lexical meaning dissolves entirely; what remains is a phonetically memorable pattern. The arrangement reinforces this effect: in the bar leading into the hook, most instruments drop out, leaving only sparse percussion and the vocal line. This sudden textural thinning creates a moment of focused anticipation, which is released when the full instrumentation crashes back in immediately after the hook. The resulting contrast between near-emptiness and saturation heightens the hook’s memorability and likely contributes to sustaining viewer retention, a key metric in Douyin’s algorithmic evaluation cycle.

Significantly, the production appears to favor the mid-to-high frequency range (roughly 2 kHz–5 kHz), which suits smartphone speakers’ typical output characteristics and aligns with the human ear’s heightened sensitivity in this band. This helps the track cut through ambient noise and capture attention within the critical first two seconds of autoplay—a technological precondition for surviving the seed-pool screening described in Section 2.1.

Paired with the visually simple, highly imitable “Ear-Covering Dance” and the track’s lighthearted mood, these sonic features coalesced into a package that lent itself to widespread imitation and resharing. By the time of this study, the “See Tình” topic on Douyin had accumulated over 900 million views.

3.2 “Зая” (Russia)

Released in 2017 by the Russian group Bambinton, “Зая” operates at 123 BPM—a medium-to-fast tempo that sits comfortably within the optimal range for dance-driven short-video content. The percussion follows a steady pattern in which the kick drum falls on the first and third beats and the snare and electronic claps on the second and fourth, producing a reliable “thump-clap-thump-clap” pulse. This regularity is functionally significant for user-generated content: every strong beat provides a dependable anchor for physical movement, allowing users to synchronize their actions without needing to edit precisely to an irregular rhythm.

The word “Зая”—a Russian term of endearment roughly equivalent to “sweetheart”—is repeatedly placed on these strong beats. For non-Russian-speaking listeners, its lexical meaning is entirely inaccessible; what registers is its sound: a soft, breathy texture that functions independently of semantics. The vocal delivery reinforces this effect. The singers use a close-miked, conversational tone, while the instrumental backing is pulled back in the mix, keeping the voices forward and unadorned. The result is a whispered proximity—as if the voice is directed not at a crowd but at a single listener holding a phone. This intimacy is highly com-

patible with the solitary, headphone-based mode of short-video consumption, and it directly activates the emotional resonance dimension of this study's analytical framework.

Melodically, the chorus is strikingly constrained: it revolves almost entirely around D4 and spans no more than a perfect fifth. This narrow range makes the vocal line easy to internalize and reproduce after a single hearing. More importantly, the melody's brevity aligns precisely with the loop-based logic of short-video playback, where a 15-second hook is replayed rather than a full song being performed. A melody that can cycle seamlessly within this window is likely to sustain high completion rates—a key signal in Douyin's algorithmic recommendation system.

To accompany the track, users devised a distinct dance pattern characterized by light, alternating footwork timed to the percussion accents. The dance is visually distinctive yet technically undemanding, lowering the threshold for imitation while still offering room for personal variation. Cumulatively, “Зая”-related topics on Douyin have amassed over 280 million views.

3.3 “Whatever”(Thailand)

A fusion of Thai pop and R&B, “Whatever” is built on a simple chord progression in C major—primarily I, IV, V, and VI—which produces a bright, open, and immediately accessible sound. This harmonic straightforwardness means the song demands no effort from the listener to follow its melodic contour, lowering the cognitive threshold for engagement from the very first seconds of exposure.

The instrumental arrangement is clearly layered in three distinct registers. The foundation rests on arpeggiated guitar chords, each note plucked separately so that the texture stays crisp and uncluttered. The middle layer features carefully restrained drum patterns, with the kick and snare processed softly so that the rhythm section supports rather than competes. The top layer introduces luminous electronic textures via synth pads that fill the mid-to-high frequency range, helping the track cut through the limited output of smartphone speakers. The contrast between the bright electronics and the warm acoustic guitar gives the song an immediately recognizable sonic signature—an advantage in a feed where attention must be captured within seconds.

The tempo sits at roughly 70-80 BPM, close to a resting heart rate. Rather than driving forward, this pace creates a gentle, breathing rhythm. The drum pattern reinforces this effect by emphasizing off-beats—placing accents on the second and fourth beats of each bar—producing a subtle swaying sensation rather than an insistent dance pulse. This rhythmic profile is crucial for understanding the song's dissemination pattern on Douyin: the tempo is too slow for full-body choreography, but ideally suited to the hand gestures and upper-body movements that became the song's signature visual template. The music thus directly shaped the form of user participation that followed.

The singer's vocal timbre is highly distinctive. She places her voice far forward in the mix and relies extensively on a breathy delivery, producing the sensation of being whispered to at close range. For listeners who do not understand Thai, the semantic content of the lyrics is entirely opaque. What registers instead is the sheer intimacy of the vocal texture itself—a tender, breathy closeness that requires no translation. This emotional immediacy, conveyed through vocal timbre alone, closely parallels the dynamic observed in “Зая”: two different languages, two distinct vocal styles, yet both activate the same mechanism of non-verbal emotional resonance.

Cumulatively, the “Whatever” topic on Douyin has amassed over 2.28 billion views.

3.4 “Montagem Nada Tropica” (Spanish)

As a fusion of funk and electronic music, “Montagem Nada Tropica” circulates on short-video platforms primarily in a sped-up version. The original tempo of around 100 BPM is accelerated to roughly 134 BPM. The slower original offers a fuller, warmer texture, while the sped-up version sharpens the attack of every sonic element—tightening the rhythm, raising the vocal pitch, and intensifying the overall sense of urgen-



cy-making it better suited to the rapid-fire autoplay environment of Douyin's feed.

The track is built almost entirely from electronic synthesis. The low-frequency foundation comes from synth bass and plucked bass, while a steady kick drum anchors the rhythm. This clean frequency separation-deep lows handled by synthesized bass, clear highs occupied by vocals and melodic lines-ensures that the main hook remains distinct even when played through smartphone speakers. The overall texture is lean, precise, and uncluttered, producing a futuristic, cyberspace-like atmosphere that distinguishes it from the warmer, more organic timbres of the other three cases.

The emotional profile of the track derives from its sonic construction: a powerful, driving beat, dense instrumental layering, and a rapidly ascending melodic contour combine to generate a strong sense of forward rush and intensity. Where "See Tinh" evokes playfulness and "Зая" and "Whatever" create intimacy, "Montagem Nada Tropicana" offers exhilaration. Crucially, this emotional quality is conveyed through sheer sonic energy and does not depend on listeners' understanding of the lyrics, directly activating the emotional resonance dimension of this study's analytical framework.

The visual template that emerged around the track-dubbed the "Sachi Shake" by Douyin creators-is built around rapid, precisely timed movements synchronized to the track's pronounced rhythmic accents. The fast tempo means that the window for executing each movement is compressed, which makes accurate synchronization more visually rewarding: when a creator locks their motion to the beat, the resulting "beat-matching" effect delivers a strong sense of visual satisfaction. This tight coupling of rhythm and movement, amplified by the track's already high-energy sonic profile, further drove imitation and sharing.

Cumulatively, related Douyin topics have amassed over 260 million views.

4. Mechanism of Viral Spread

4.1 Algorithm-Driven Dissemination

The dissemination of all four songs began with Douyin's tiered recommendation algorithm. When users upload short videos containing these songs, the platform evaluates content appeal through quantitative engagement metrics-watch time, likes, comments, saves, and shares. Videos that generate strong initial signals are then pushed to progressively larger user pools, initiating a self-reinforcing cycle in which positive feedback and expanded distribution feed into each other. This is the decentralized mechanism described in Section 2.1, and every case in this study followed this same trajectory.

What is striking across the four cases is that none of them depended on listeners' comprehension of the lyrics to generate the initial engagement that triggered broader distribution.[17] The Vietnamese "ding-ding-dong-dong" registered as playful onomatopoeia, not as meaningful speech. The Russian "Зая" was received as an intimate vocal texture rather than a word with a dictionary definition. The Thai chorus conveyed tenderness through breathy timbre alone. The accelerated track delivered its impact through sheer sonic density. In each instance, the algorithm's evaluation was driven by how users responded to sound, not by what they understood from language.

The musical structures analyzed in Chapter 3 help explain why these responses were consistently strong. Each song contains a hook that appears designed-whether intentionally or through production choices that happened to align with short-video consumption-to capture attention within the critical first seconds of autoplay. The build-and-release arrangement of "See Tinh", the whispered vocal proximity of "Зая", the gestural pulse of "Whatever", and the high-energy rush of "Montagem Nada Tropicana" all created conditions conducive to high viewer retention. A high completion rate is among the most influential signals in Douyin's recommendation logic. When users additionally replayed the same video, this behavior may have further signaled content quality to the algorithm, amplifying exposure still further.

In this way, the platform's algorithm functioned as the first bridge enabling regional languages songs to cross linguistic barriers. It reduced content competition to a data-driven level where neither promotional re-

sources nor linguistic origins solely determined a song's reach. The algorithm did not recognize these songs as "foreign-language" or "niche"; it recognized them only as content that users watched, liked, and shared. This indifference to language, built into the very architecture of the recommendation system, was the pre-condition that made all subsequent mechanisms-sonic hooking, template-driven participation, and emotional resonance-possible on a mass scale.

4.2 Sonic Hook Mechanism

If the algorithm provides the channel through which regional languages songs are discovered, the sonic hook determines whether they are remembered. Across all four cases examined in this study, the musical segment that circulated most widely was not a full song, nor a verse chosen for its lyrical meaning, but a brief, self-contained snippet engineered-whether by design or by accident of production-to grab attention within seconds and lodge itself in memory. There are three shared characteristics of these hooks help explain why they functioned so effectively in the short-video environment, each corresponding to a distinct facet of the "acoustic appeal" dimension set out in this study's analytical framework.

The first is semantic detachability. In every case, the hook's communicative power did not depend on the listener understanding what the words meant.[18] The Vietnamese "ding-ding-dong-dong" operated as pure onomatopoeia-its rhythmic, phonetic shape was the message, not any lexical content it may carry for a native speaker. The Russian "Зая", a term of endearment in its original context, registered for non-Russian listeners as a soft, breathy vocal texture rather than a word with a dictionary definition. The Thai chorus, delivered through a close-miked, breathy vocal style, conveyed tenderness through timbre alone, bypassing the semantic layer entirely. "Montagem Nada Tropica", once accelerated, compressed its syllables to the point where they became rhythmic components within a larger sonic rush rather than carriers of discrete meaning. In each of these instances, the language barrier did not weaken the hook; it stripped away the cognitive work of linguistic processing and left behind something more immediate-sound as sensory experience, accessible to any listener regardless of language background.

The second is structural compatibility with the short-video playback loop. Douyin videos are measured in seconds, not minutes. A hook that requires buildup, narrative development, or extended listening to make its impact is structurally unsuited to this environment. The four hooks in this study, by contrast, are all conspicuously short, self-contained, and repeatable. "See Tinh" delivers its four-syllable chant within a single bar. "Зая" places its one-word hook squarely on the strong beats of a steady percussion pattern. "Whatever" confines its chorus melody to a narrow interval range-no more than a perfect fifth-making it reproducible almost instantly. "Montagem Nada Tropica", in its sped-up form, packs its entire hook into a dense burst of electronic energy that hits with full force from the first beat. Each of these fragments can play out completely within a fifteen-second window, cycle multiple times, and still feel musically coherent rather than truncated. This structural fit between hook length and video duration contributes to sustained viewer retention: a user who reaches the end of a clip has, by definition, heard the full hook, and the seamless loop invites replay rather than signaling an endpoint.

The third is production-level adaptation to mobile playback conditions. Smartphone speakers, the default listening device for short-video consumption, have well-documented limitations in reproducing low frequencies. The hooks in this study all appear-whether through conscious mastering decisions or stylistic convention-to favor the frequency range where these devices perform best and where the human ear is most sensitive. "See Tinh" exhibits a pronounced presence in the 2 kHz-5 kHz band, helping it cut through ambient noise. The three-layer arrangement of "Whatever" places luminous synth pads in the mid-to-high register above a restrained rhythm section, ensuring the hook's textural signature remains clear even on small speakers. "Зая" takes an alternative but equally effective approach: by pulling back the instrumental mix and pushing the vocals far forward, it creates an intimate, close-miked effect that is ideally suited to headphone listening, the other dominant mode of short-video consumption. The accelerated version of "Montagem Nada Tropica" sharpens the attack of every element, raising the overall spectral energy in a way that



translates forcefully even through limited hardware. Across these varied strategies, the result is the same: the hook sounds distinct, legible, and attention-commanding under the actual listening conditions in which short-video users encounter it.

In conclusion, these three characteristics-semantic detachability, structural compatibility with the loop, and mobile-optimized production define the sonic hook as the second link in the dissemination chain identified by this study. The algorithm, as described in Section 4.1, opens the door; the hook determines whether the user stays, finishes the clip, replays it, and remembers it. Without this retention, no further mechanism-no visual template, no participatory challenge, no emotional bonding can be activated.

4.3 Participatory Templates

A song that is merely catchy can be listened to; a song that is paired with a replicable visual action can be participated in. Across all four cases, the transition from passive reception to active engagement occurred through the same mechanism: a core musical hook was extracted and bound to a distinctive, easy-to-imitate movement template. This binding transformed the song from a piece of audio into a participatory package-compact, shareable, and open to variation.

The pattern was remarkably consistent. In “See Tinh”, the onomatopoeic “ding-ding-dong-dong” hook was matched with the “Ear-Covering Dance,” a simple upper-body movement that required no dance training to replicate. In “Зая”, the steady “thump-clap-thump-clap” pulse and the repeated endearment were paired with a light, alternating footwork pattern that users could perform in place. In “Whatever”, the gentle, off-beat sway of the chorus was linked to a hand-gesture challenge involving fluid upper-body and arm movements. In “Montagem Nada Tropica”, the accelerated electronic rush was coupled with the “Sachi Shake”, a rapid, precisely timed sequence of motions designed to lock onto the track’s pronounced beats. In every instance, the visual template was not an afterthought appended to the music but an integral part of what made the song spread.[19]

What made these templates so effective as vehicles of dissemination can be understood on three levels.

Firstly, the templates drastically reduced the creative threshold for participation. The user did not need to invent a dance, choreograph a sequence, or conceptualize an original visual concept. A ready-made format was provided-a set of movements that anyone could observe, memorize after one or two viewings, and reproduce with a phone camera. The cognitive effort required was minimal, while the sense of achievement upon completing the imitation was immediate. As described in Section 2.2, Douyin’s editing tools and topic-based challenge system further supported this process by allowing users to extract the relevant music snippet with a single tap and place their own video alongside countless others under the same hashtag.

Secondly, the templates enabled personal variation within a shared framework. While the basic movements were fixed enough to be recognizable, they were also loose enough to accommodate individual interpretation. Users could perform the “Ear-Covering Dance” with different facial expressions, settings, or costumes. The footwork for “Зая” could be executed with varying levels of energy and style. The hand-gesture challenge for “Whatever” invited subtle differences in how each person moved their arms and fingers. The “Sachi Shake” could be intensified or softened depending on the creator’s physicality. This combination of standardization and flexibility meant that participation did not feel like mechanical copying; it allowed for a sense of personal ownership while still contributing to a collective, recognizable trend.

Finally, the templates were inherently social. Each video posted under a song’s topic hashtag was simultaneously an act of individual expression and a contribution to a growing collective phenomenon. Users who scrolled through the topic page could see thousands of others performing the same template, which created a dual motivation: the desire to join a visible community and the reassurance that the barrier to entry was low, since so many others had already done it. Analysis of the comment sections under the most widely circulated videos for each song supports this reading. Comments repeatedly expressed the desire to participate, the sense of being part of a collective trend, and the wish to showcase one’s own version. These discursive

patterns indicate that participation satisfied multiple needs at once: the need for social validation through engagement with trending topics, the need for entertainment through playful imitation, the need for virtual connection with other users and with the original creators via duet and remake features, and the need for self-expression through creative personalization.

The participatory template was thus the mechanism through which the sonic hook's retention power was converted into sustained, large-scale user engagement. The hook made the song memorable; the template gave users something to do with that memory. Without this step, the song would remain an auditory experience passively consumed. With it, the song became a site of active, visible, and socially rewarding participation—a condition that algorithmically amplified it still further, as each new video fed back into the engagement metrics that drove continued recommendation. In the chain of dissemination, the template was the link that turned listeners into disseminators.

5. Conclusion

Through an analysis of the dissemination pathways of four regional languages songs on Douyin, this paper draws the following conclusions.

In essence, the widespread popularity of these songs embodies a form of decentralized dissemination. In traditional music distribution, understanding the specific meaning of lyrics is often a prerequisite for listener engagement; however, within Douyin's ecosystem, this premise has been fundamentally overturned. The fact that listeners cannot understand the lyrics does not necessarily impede their engagement; on the contrary, under the conditions examined in this study, the suspension of semantic comprehension can redirect attention toward the sonic surface. When the semantic function of linguistic symbols recedes, their timbral properties come to the fore, shifting listeners' focus from "what this song is singing about" to "how this song sounds".

The core dissemination mechanism can be summarized as the synergy of four interrelated elements: algorithm, sonic hook, participatory template, and emotional resonance.

The algorithm provides the channel, determining how songs are discovered. Douyin's decentralized recommendation mechanism grants equal exposure opportunities to any content with potential, evaluating videos purely through engagement metrics rather than by the creator's identity or the language of the lyrics.

The sonic hook serves as the bridge between discovery and retention. Across all four cases, the musical segment that circulated most widely was a brief, self-contained snippet in which semantic content was either absent or irrelevant to non-native listeners. These hooks shared three characteristics—semantic detachability, structural compatibility with the short-video loop, and production-level adaptation to mobile playback—that together ensured they could capture attention within seconds and sustain high completion rates.

The participatory template provides the form, shaping how listeners interact. By binding a core musical snippet to a distinctive, easy-to-imitate visual action, each song was transformed from something merely pleasant to listen to into something people could actively engage with. This binding drastically lowered the creative threshold for participation while still allowing room for personal variation, turning passive recipients into active disseminators.

Emotional resonance forms the deepest layer, explaining why these songs move their listeners.[20] The four songs span distinct emotional registers—playfulness, intimacy, warmth, and exhilaration—yet all four reached their audiences not through the semantic content of the lyrics but through the sonic surface itself: the texture of the voice, the speed of the beat, the density of the arrangement. Listeners who understood not a single word nonetheless described these songs in terms of how they felt, not what they said.

These four elements are indispensable and mutually reinforcing. Without the algorithm, the songs would remain undiscovered. Without the sonic hook, initial exposure would not convert into sustained attention. Without the participatory template, retention would not translate into active engagement. And without emotional resonance, none of the preceding mechanisms could generate the voluntary, sustained, and socially



contagious participation that defines the phenomenon this study set out to explain.

From the perspective of the music industry, this phenomenon offers several practical insights. Music production wishing to optimize for short-video circulation may benefit from deliberate attention to the design of a hook, whether a catchy vocal phrase, a clean instrumental interlude, or a simple, recurring lyric. Its defining traits are brevity, memorability, and imitability: the opening seconds of a song often determine its fate on short-video platforms. During the arrangement stage, composer should consider mobile-speaker playback quality, anticipate which segments users are likely to clip as video soundtracks, and assess whether the music lends itself to simple, replicable movements.

From the perspective of cross-cultural communication, this phenomenon also opens new avenues for reflection.[21] In the short-video era, the universality of emotions and embodied behaviors may matter more than linguistic commonality. Traditional cross-cultural communication often depends on translation and the accurate understanding of meaning, but the dissemination pathways observed on Douyin demonstrate an alternative possibility: conveying emotions directly through sound and movement. As the four cases in this study suggest, when the linguistic barrier suspends semantic processing, listeners' attention can be redirected toward the sonic surface—the grain of the voice, the pulse of the rhythm, the contour of the melody—enabling an unmediated encounter with the emotional substance of the music itself. This mode of communication requires neither translation nor explanation to resonate across cultural boundaries.

Finally, it should be noted that the success of these four songs, while following a shared underlying logic, does not imply any fixed formula for widespread dissemination. Each song possesses its own distinct sonic character, and every visual template originated from a specific creative act within a particular user community. Algorithms can amplify content, but they cannot create it; templates can facilitate participation, but they cannot replace the spontaneous appeal of the music; and emotional resonance, however systematically analyzed, ultimately depends on something that cannot be engineered—the capacity of sound itself, stripped of semantic meaning, to evoke a genuine emotional response.[22]

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Author Contributions

The author conducted all research and wrote the manuscript.

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

- [1] Xu Zhu. The Communication Structure, Characteristics, and Value of Mobile Short Videos [J]. Journalism Lover, 2019, (12): 30-32.
- [2] Huang Xiaoyin, Qiu Zihao. Technology Empowerment and Emotional Interaction: A Study on Visualized Music Communication on the Douyin Platform [J]. Journal of Southwest Minzu University (Humanities and Social Sciences Edition), 2019, 40(08): 156-161.
- [3] Duan Peng. On the Visual Turn of Music Communication in the Era of Short Videos [J]. Modern Publishing, 2022, (01): 45-49.
- [4] Yang Ting, Zhou Ran. An Analysis of the Creative Characteristics of Short-Video Songs within the Online Art Ecology [J]. Art Evaluation, 2022, (13): 13-16.
- [5] Li Dingdian. The Hidden Hand: Auditory Dominance in Short-Video Communication [J]. New Media

- Research, 2022, 8(12): 52-56.
- [6] Liu Dequn. Research on the Operational Logic, Value Concealment, and Collaborative Guidance Paths of Online Short Videos from the Perspective of Intelligent Algorithms [J]. Journal of Beijing University of Posts and Telecommunications (Social Sciences Edition), 2023, 25(02): 48-55.
 - [7] Wang Wenyi. From “Poaching” to Participation: The Identity Shift of Short-Video Secondary Creators [J]. All-Media Exploration, 2023, (12): 132-133.
 - [8] Yu Jia, Zhang Jing. Technology Empowerment and Value Transmission: Emotional Resonance in the Soundtracks of Documentary-Style Short Videos [J]. Music Education and Creation, 2023, (04): 26-31.
 - [9] Liu Jia. An Invisible Force: Sound Practices in Short-Video Communication [J]. China Television, 2024, (01): 92-96.
 - [10] Zhang Kaishuo, Zhang Jingyi. An Analysis of Song Reviews and Music Communication Strategies from a Cross-Cultural Perspective [J]. China Nationalities Expo, 2024, (18): 113-115.
 - [11] Li Xingying. An Analysis of the Intelligent Algorithm Mechanism in Douyin Short Videos [J]. Electronic Technology, 2022, 51(09): 204-205.
 - [12] Wu Bin, Zhang Zhaoxuan. A Study on Users’ Perception of Algorithmic Recommendations on Short-Video Platforms: Taking Douyin as an Example [J]. Journal of Hangzhou Dianzi University (Social Sciences Edition), 2024, 20(02): 27-35.
 - [13] Hua Ruiyun. Research on the Communication Mechanism and Countermeasures for Background Music in Short Videos [J]. Editors Monthly, 2023, (01): 111-115.
 - [14] Wang Xuecheng, Yang Haochen. The “Constellation” of Simulacra: A Critical Interpretation of Short-Video Secondary Creation [J]. Modern Communication (Journal of Communication University of China), 2022, 44(03): 151-156.
 - [15] Wen Jing. An Exploration of Operational Strategies for Music Visualization Accounts on Short-Video Platforms: Taking Douyin as an Example [J]. China Media Technology, 2023, (04): 95-99.
 - [16] Li Zhen. An Exploration of How Sound Elements in Short Videos Capture Audience Attention: Taking the Douyin Platform as an Example [J]. Voice & Screen World, 2024, (03): 96-98.
 - [17] Guo Xuejiao. An Exploration of Content Production and Anomie in Music Short Videos [J]. New Media Research, 2018, 4(11): 16-17.
 - [18] Huang Wei. An Exploration of Communication Strategies for Music-Based Short Videos in the New Media Era [J]. China Media Technology, 2024, (12): 60-63.
 - [19] Han Lu. The Construction of a Music Communication System in a Converged Media Environment: Taking the “Douyin” Short-Video Platform as an Example [J]. Northern Music, 2020, (11): 240-241.
 - [20] Li Jiayi, Liu Peng. Domestication and Reverse Domestication: An Empirical Study on the Algorithmic Recommendation Phenomenon in Douyin Live Streaming [J]. China Media Technology, 2023, (05): 72-75.
 - [21] Xiong Jieni. Research on Short Videos and Cross-Cultural Communication: Taking TikTok as an Example [J]. Journalist Observation, 2024, (23): 86-88.
 - [22] Liu Keyi. Identification of Copyright Infringement on Short-Video Platforms under Algorithmic Recommendation Technology [J]. Pioneering with Science & Technology Monthly, 2024, 37(02): 166-170.

