How do we process words in music? A qualitative study with Gen Z listeners

Mauro Cavallone, University Parthenope of Naples, Department of Management and Quantitative Studies, mauro.cavallone@uniparthenope.it
Luca Nobis, CPM Music Institute, Academic Director of Courses and Head of the Pop Rock Guitar Department, lnobis@cpm.it
Andrea Pozzi, University of Bergamo, Department of Foreign Languages, Cultures and Literatures, andrea.pozzi@unibg.it
Rocco Palumbo, University of Tor Vergata of Rome, Department of Management, rocco.palumbo@uniroma2.it

Abstract: "Prisencolinensinainciusol ol rait": this opening line from Adriano Celentano's 1972 song — composed of invented, English-sounding nonsense words — became an international hit. If a song without semantic content can move millions, what does this reveal about how we process music? This study examines what captures listeners' attention in a song and whether musical expertise shapes that focus. Two Gen Z groups — non-musicians and students pursuing musical training — listened to the same song across three progressive modalities: audio only, audio with English lyrics, and audio with lyrics and Italian translation. Written responses to four open-ended prompts trace what captures attention, what emotions are felt, what meaning is perceived, and what is retained. Preliminary findings suggest two distinct profiles. Non-musicians show early and stable semantic engagement. Musically trained listeners attend to structural elements first, with meaning emerging only upon explicit textual input. In both groups, perceived meaning converges fully by the final phase.

Keywords: Music, lyrics, emotion, selective attention, perception.

1. Introduction

Music plays a dominant role in everyday life. It features prominently in advertising, film, public events, and political messaging to engage audiences. While music and lyrics are believed to work together to produce a combined effect, research suggests that musical and verbal elements are processed differently—at varying speeds and levels of consciousness (Senn et al., 2023; Bourne, 2025). An interesting example is Adriano Celentano's 1972 song, where the opening line is "Prisencolinensinainciusol ol rait"—a string of English-like words with no actual meaning. This raises a question: if a nonsensical international hit can captivate millions and evoke emotion, what does this suggest about how people perceive and interpret music?

This exploratory study assumes that individuals do not process musical and verbal content simultaneously when listening to a song. Instead, they focus first on elements that are immediately accessible, which can depend on the listening mode and their knowledge of the music (Deutsch, 1999; Juslin & Laukka, 2004). To explore this, two groups of Gen Z participants—one with formal musical training and one without—listened to the same song in three ways: audio-only, audio with English lyrics, and audio with Italian translation. Their written responses to four open-ended questions—what attracts attention, evokes emotion, is interpreted as

meaning, and is ultimately remembered—were analysed both quantitatively and qualitatively. Differences between the groups were also explored to understand how perceptions changed according to listening mode and participants' familiarity with music.

2. Conceptual framework

Music is an ever-present part of daily life. People listen to it while commuting, working, shopping, and socialising — often unintentionally and without focusing on it. Studies on everyday music listening show that, for most people and in most settings, its main role is emotional rather than semantic: music helps regulate mood, maintain energy, or offer comfort, rather than being used to understand or interpret lyrics (Sloboda & O'Neill, 2001). However, scholars have emphasised that the contact between people and music goes beyond emotions, shaping the cognitions and the mental processes of individuals (Taylor, 1982). While there is agreement that music is central in improving the way people deal with everyday life dynamics and is a resource for well-being, creativity, and performativity (*e.g.*, Stige, 2007; Tracey, 2011; Odena, 2012), there is less consensus on how people actually make sense out of their exposure to music (Reybrouck, 2021) and how they react to it (Reybrouck et al., 2024). This leads to a central question for this study: if responses to music derive from a multiplicity of triggers and do not necessarily imply a full understanding of the lyrics, what exactly do listeners react to — and in what way?

Studies on how we perceive musical sound generate a partial understanding of this topic. Deutsch (1999) highlighted that listeners do not perceive music as a single, unified entity. Instead, they organise it into perceptual configurations, grouping and separating its elements based on principles that are not entirely within conscious awareness. Different listeners may focus on various aspects such as melody, rhythm, timbre, or voice within the same musical passage. Furthermore, listeners' motives and the context in which they come into contact with the music shape how they extrapolate meaning and emotions from music (Juslin & Laukka, 2004). The mode of listening and the available knowledge about what is listened to add to the complexity of the process that enables people to obtain cues and insights from music (Bourne, 2025). Finally, of equal relevance, source sensitivity and the complexity of the piece listened to have an important effect on how people react to music (Senn et al., 2023; Chilvers et al., 2024), increasing the difficulty of predicting how senses and emotions are obtained from the music experience. Based on these considerations, it is possible to claim that perceptions differ significantly from the written score or the composer's intention, especially when there is unawareness of the cultural attributes and the characteristics of the artist (Kiernan et al., 2022). This selective and variable perception indicates that listening is not a passive reception of an objective signal but an active, constructive process influenced by prior experience, attention, and cultural background.

Juslin and Västfjäll (2008) explored how music triggers emotions through various psychological mechanisms functioning at different levels of awareness. Some mechanisms, like rhythmic entrainment and brainstem reflexes, are quick, automatic, and largely unaffected by the music's meaning. Others, such as episodic memory and cognitive appraisal, are slower and more influenced by the listener's personal history and interpretive

effort. This view suggests that emotional reactions to music are complex and layered, with different parts of the musical signal influencing feelings at different speeds and levels of conscious engagement.

Building on these observations, the present study proposes that when listeners engage with a song — a form that combines musical and verbal content — the two dimensions are not processed simultaneously or to the same depth. The musical layer tends to engage listeners first and most immediately, generating emotional and cognitive responses that may precede and operate independently of any elaboration of the lyrical content. Verbal meaning, when accessible, enters the listening experience later, and cognitively shapes it in ways that depend on the listener's linguistic competence, prior exposure, and attentional orientation. We argue that the emotional and cognitive dimensions of music perception are processed selectively and asymmetrically, with the musical layer capturing attention before — and at times independently of — the verbal layer. This paper offers preliminary, exploratory observations in support of this view and discusses their implications for how music functions as a communicative medium.

3. Methodology

This study uses a qualitative approach to explore how listeners interpret the musical and verbal elements of a song during different stages of listening. Data were gathered via a structured questionnaire from two groups of Gen Z participants in June 2026: master's students at the University of Bergamo "UNIBG" (sample, n=12; data collection: 15 June), who have no formal musical training, and students at the CPM Music Institute in Milan "CPM" (sample, n=6; 17 June), who possess formal musical education. Musical expertise within the CPM group was inferred from their academic background rather than directly assessed.

Participants were invited to listen to a short excerpt of the same song — "Walk a Mile in My Shoes" (the version performed by Elvis Presley in 1970[1]) — across three progressive stages: audio only (Phase A), audio with English lyrics (Phase B), and audio with an Italian translation of lyrics (Phase C). After each phase, participants responded in writing to four open-ended questions: what captured their attention (Q1), what emotions the song conveyed (Q2), what they perceived as its meaning or message (Q3), and what remained most salient at the end of the listening experience (Q4).

Data analysis relied on a coding procedure based on a thematic categorisation scheme developed inductively from the data (Braun & Clarke, 2006) around 3 macro-categories and 15 categories:

- Musical elements: Rhythm, Melody, Specific instruments, Voice, Lyrics/words, Music-text contrast.
- Emotional responses: Positivity, Energy, Melancholy, Contrast, Value engagement/empathy.
- Residual categories: No response, Incorrect, Vague/generic, Translation contrast.

A single category was assigned to each response based on the predominant theme. Inter-rater reliability was evaluated through independent coding by another researcher. The findings are presented as descriptive frequencies and qualitative observations — both for the total sample and by group — and no inferential statistical tests were conducted due to the exploratory scope of the study and the small sample size (Patton,

[1] The song was written by Joe South in 1970 and performed by Elvis Presley in the same year. The short excerpt comes from the video available at: https://www.youtube.com/watch?v=fUKPcStDjwQ&list=RDfUKPcStDjwQ&start_radio=1 (accessed: 14 June 2026).

2002). The results are organised around two analytical dimensions: the role of selective processing across listening phases (Section 4.1), and the effect of musical expertise on elaboration profiles (Section 4.2). These observations serve to generate hypotheses for future research with more robust designs.

4. Findings

4.1 What captures listeners' attention and what remains

The comparison of what captured listeners' attention (Q1) and what they remembered (Q4) across the three listening phases illustrates how musical and verbal elements are processed. Tables 1 and 2 display the percentage for each category at each stage.

Table 1: What captured listeners' attention: percentage values across phases. All respondents.

Macro-category	Category	Phase A	Phase B	Phase C
Musical elements	Rhythm	50%	22%	28%
	Melody	11%	22%	6%
	Specific instruments	17%	6%	6%
	Voice	17%	17%	11%
	Lyrics/words	0%	6%	22%
	Music-text contrast	6%	17%	17%
Emotional/semantic	Value engagement/empathy	0%	11%	0%
Other	Vague/generic / No response	0%	0%	11%

Table 2: What stayed with listeners most (Q4): distribution across phases. All respondents.

Macro-category	Category	Phase A	Phase B	Phase C
Musical elements	Rhythm	44%	17%	0%
	Specific instruments	17%	0%	0%
	Voice	22%	0%	0%
	Melody	0%	11%	0%
	Music-text contrast	0%	6%	0%
	Lyrics/words	0%	28%	50%
Emotional/semantic	Value engagement/empathy	6%	17%	39%
	Positivity	6%	11%	6%
Other	Vague/generic / No response	6%	11%	6%

In Phase A — which involves only audio — listeners immediately focus on musical components, with "Rhythm" being the most dominant at 50%, followed by “Voice” and “Specific instruments” at 17% each. These same elements are also what listeners remember most at the end of the session — "Rhythm” (44%), “Voice” (22%), and “Specific instruments” (17%) — together accounting for over 80% of the responses. The

perceived meaning of the lyrics and emotional aspects are either not present or mentioned only a few times. During Phase A, the musical layer is what captures and retains attention regardless of any verbal content, as confirmed by participants' own words:

> *"Il ritmo travolgente"* [The overwhelming rhythm] (UNIBG 3, Phase A)
> *"Accompagnamento coinvolgente e ritmato"* [An engaging and rhythmic accompaniment] (UNIBG 10, Phase A)
> *"Ritmo incalzante"* [A driving rhythm] (UNIBG 5 and 6, Phase A)
> "La presenza di tanti strumenti diversi, DING acuto, basso e chitarra galoppano, batteria super grooovie" [The presence of so many different instruments — a sharp DING, bass and guitar galloping, super groovy drums] (CPM 13, Phase A)
> "Il rullante e l'intro di chitarra" [The snare drum and the guitar intro] (CPM 15, Phase A)
> "Capacità del cantante di creare uno scenario ritmico con la voce" [The singer's ability to create a rhythmic landscape with his voice] (CPM 17, Phase A)

Listeners' focus changes when English lyrics are introduced in Phase B. Two new categories emerge in Q1: Music-text contrast (17%) and Value engagement/empathy (11%). "Lyrics/words" is what listeners remember most at the end of the session (28%), suggesting that lyrics can generate a semantic response without translation. The noticeable contrast between music and lyrics seems clear from how listeners responded:

> *"Il contrasto tra testo e spensieratezza della musica"* [The contrast between the lyrics and the carefree quality of the music] (UNIBG 2, Phase B)
> *"Il ritmo stravolgente in contrasto con il messaggio diretto a chi critica il cantante"* [The overwhelming rhythm in contrast with the message directed at those who criticise the singer] (UNIBG 5, Phase B)
> *"Stupore, perché senza il testo pensavo fosse tutt'altro"* [Surprise, because without the lyrics I thought it was something completely different] (UNIBG 1, Phase B)

The most striking pattern emerges in Phase C, when participants listened to the audio with Italian translation of lyrics. In Q1, "Rhythm" remains the most cited attentional element (28%), and "Lyrics/words" rises to 22%. The picture changes in Q4: "Rhythm" drops to 0%, while "Lyrics/words" rises to 50% and Value engagement/empathy to 39%. This suggests that understanding verbal content shifted listeners from a rhythmically focused experience to one more defined by semantic and emotional richness. Participants' responses illustrate this shift.

> *"Il fatto che venga trattato un tema così importante con questo tipo di musica"* [The fact that such an important theme is addressed through this kind of music] (UNIBG 2, Phase C)
> *"Non avrei mai associato il ritmo col tema"* [I would never have associated the rhythm with this theme] (UNIBG 1, Phase C — retained from Phase B)
> *"In italiano il testo ha ancora più significato"* [In Italian the lyrics have even more meaning] (UNIBG 9, Phase C)

4.2 Does musical expertise make a difference? Two distinct elaboration profiles

Between-group comparisons allow us to explore whether musical familiarity and knowledge change listeners' perceptions. Table 3 shows how attentional responses (Q1) are distributed across groups in each phase, whereas Table 4 describes the evolution of perceived meaning (Q3) for each group.

Table 3: What captured listeners' attention (Q1): percentage values by group and phase.

Category	UNIBG Phase A	CPM Phase A	UNIBG Phase B	CPM Phase B	UNIBG Phase C	CPM Phase C
Rhythm	67%	17%	25%	17%	33%	17%
Melody	8%	17%	17%	33%	8%	0%
Specific instruments	8%	33%	0%	17%	0%	17%
Voice	8%	33%	8%	33%	0%	33%
Lyrics/words	0%	0%	8%	0%	33%	0%
Music-text contrast	8%	0%	25%	0%	17%	17%
Value engagement/empathy	0%	0%	17%	0%	0%	0%
Other	0%	0%	0%	0%	8%	17%

Table 4: Perceived meaning (Q3): summary of trajectories by group.

	Phase A	Phase B	Phase C
UNIBG (n=12)	Value engagement/empathy 50% — Other 50%	Value engagement/empathy 100%	Value engagement /empathy 100%
CPM (n=6)	Value engagement /empathy 0% — No response/Incorrect 50%	Value engagement/empathy 83% — Incorrect 17%	Value engagement /empathy 100%
Total (n=18)	Value engagement/empathy 33%	Value engagement /empathy 94%	Value engagement /empathy 100%

UNIBG participants show an anticipatory-stable profile: listeners immediately perceive the song as a whole, with attention dominated by Rhythm (67%), and responses tend to be immediate and emotionally undifferentiated (Phase A). As verbal content is introduced (Phase C), attention shifts progressively towards the textual dimension: “Lyrics/words” rises to 33%, while “Rhythm” remains present (33%). Emotional responses follow a parallel trajectory, becoming progressively deeper and more articulate across phases (Table 3). The summary of trajectories (Table 4) shows that half of the group identifies the song's empathic theme in Phase A and maintains this perception across all subsequent phases. The other 50% shift their focus to “Value engagement/empathy” in subsequent stages. No UNIBG participant requires the Italian translation to access the song's meaning.

The shift towards an integrated elaboration of musical and semantic components clearly emerges in participants’ responses.

> *"Il ritmo travolgente"* [The overwhelming rhythm] (UNIBG 3, Phase A)
> *"Invita all'empatia"* [It invites empathy] (UNIBG 9, Phase A — before reading the lyrics)
> *"Vestire i panni altrui, non trattare male gli altri"* [Walk in someone else's shoes, do not treat others badly]

(UNIBG 7, Phase A — before reading the lyrics)
"Contrastanti: il suono incoraggia allegria, il messaggio quieta riflessione" [Contrasting: the sound encourages joy, the message invites quiet reflection] (UNIBG 2, Phase B)
"Il fatto che venga trattato un tema così importante con questo tipo di musica" [The fact that such an important theme is addressed through this kind of music] (UNIBG 2, Phase C)

CPM participants display a technical-evolutionary profile, with immediate attention to specific elements like "Voice" (33%) and "Specific instruments" (33%). This suggests an analytically driven listening approach that persists across all three phases. In the final session, "Voice" remains the leading category at 33%, and no CPM participant shifts attention to "Lyrics/words", even when both lyrics and translation are accessible. This technical focus influences semantic interpretation: none of the CPM participants recognises the empathic theme in Phase A (Table 3). The summary of trajectories (Table 4) shows that five out of six CPM participants align on Value engagement/empathy in Phase B, while only one does so in Phase C.

"La presenza di tanti strumenti diversi, DING acuto, basso e chitarra galoppano, batteria super grooovie" [The presence of so many different instruments — a sharp DING, bass and guitar galloping, super groovy drums] (CPM 13, Phase A)
"Il dialogo tra l'orchestrazione dei suoni e le parole, l'evoluzione della canzone, il modo in cui è espresso" [The dialogue between the orchestration and the lyrics, the song's evolution, the way it is expressed] (CPM 14, Phase B)
"Adesso conoscendo il cantante, la mia attenzione è catturata dalla probabile storia che ha ispirato il brano" [Now that I know the artist, my attention is drawn to the probable story that inspired the song] (CPM 15, Phase C)

In Phase A, one participant constructs an entirely incorrect interpretation: this response illustrates how, without linguistic comprehension and with attention directed toward technical-musical features, listeners may project generic emotional associations onto the musical stimulus rather than accessing its semantic content.

"Si riferisce ad una donna e le si riservano parole d'amore" [It is about a woman, and words of love are addressed to her] (CPM 17, Phase A)

Between-group comparisons reveal two clear profiles with one point of convergence. Both groups fully agree on "Value engagement/empathy" by Phase C, suggesting that the song's empathic theme is ultimately comprehensible to all listeners, regardless of their musical expertise—though their paths to understanding and the timing differ. For UNIBG participants, meaning emerges early and quickly becomes solid, with emotional responses intensifying over time. CPM participants, by contrast, depend on explicit verbal cues to interpret meaning, with their emotional responses staying relatively steady and closely tied to the music throughout. English proficiency might have influenced these outcomes: UNIBG participants who recognise the empathic theme in Phase A could do so partly because they understand the song's title and some lyrics. This aspect was not directly studied and should be explored further in future research.

5. Discussion and conclusion

This study aimed to examine (i) what captures listeners' attention in a song and (ii) whether musical expertise affects that focus. The findings offer a clear, though preliminary, answer to both questions.

Throughout all three listening phases, the musical component — and rhythm in particular — dominates immediate attention. In Phase A, before lyrics are heard, rhythm alone accounts for half of the responses to the first prompt (Q1: what captured your attention). At this stage, perceived meaning is vague, absent, or incorrect in most cases. As verbal content is introduced, attention diversifies, and semantic understanding begins to develop — converging on value engagement/empathy in every case by Phase C. Yet even in Phase C, rhythm remains the most cited element in Q1 (28%). The picture changes in Q4 (what stayed with you most): lyrics/words rise to 50%, value engagement/empathy to 39%, and rhythm drops to 0%. The musical layer draws initial attention; verbal content is what consolidates in memory.

Findings suggest that the relationship between music and text extends beyond a simple framing effect. Music appears to shape the process through which meaning is built, altering the sequence in which attention, emotion, and semantic content are elaborated. A contribution of this study lies in its focus on the temporal dimension of listening: the difference between groups and phases is not only in what is perceived, but in when and through what pathway meaning is reached.

The contrast between the two groups highlights this dynamic. UNIBG participants exhibit an anticipatory-stable profile: semantic perception appears in 50% of cases in Phase A — even before encountering the lyrics — and remains stable without additional input. Emotional responses grow stronger across phases, shifting from general positivity to more value-based categories. CPM participants, by contrast, display a technical-evolutionary profile: their attention from the beginning focuses on structural elements — instruments, voice, and arrangement — with semantic perception only occurring upon explicit textual input. This does not suggest a lower interpretive ability, but indicates a different attentional hierarchy. Musical expertise appears to guide attention toward the formal properties of the music, with semantic understanding developing at a later stage.

This observation raises a broader point: listeners do not process all parts of a song equally or simultaneously. The distribution of attentional resources during listening appears to be shaped by prior experience, which has implications for music's role in communication. When music is paired with a verbal message, it may not simply reinforce that message: it may alter the sequence of its elaboration or delay it, depending on the listener's background.

A further finding concerns the role of contrast. Several participants report surprise when confronted with the song's lyrical content in Phase B. The gap between the carefree musical character of the piece and the weight of its message appears to heighten engagement with the verbal content. Whether this contrast effect operates differently when music and text are more congruent remains an open question for future research.

Several limitations should be acknowledged. The elicitation instrument yielded individual written responses rather than group discussion, making this study closer to a structured survey than to a focus group in the strict sense (Morgan, 1997). The progressive listening design introduces a potential familiarity effect, though the surprise responses observed in Phase B suggest that textual input, rather than mere re-exposure, drives the shift

in perception. The sample is small and non-representative, and the findings should be treated as exploratory rather than generalisable. Musical expertise in the CPM group was inferred from participants' study path rather than directly measured, and English language proficiency — a potentially confounding variable in Phase A — was not assessed. Finally, the study relies on a single song; whether the observed patterns generalise to other genres or to songs with greater congruence between musical and verbal dimensions remains an open question. Future research should address these limitations through a design involving parallel groups, a collective discussion phase, broader and more diverse samples, and multiple song choices spanning different genres and degrees of music-text congruence. The direct measurement of musical expertise and language proficiency would also strengthen comparability. A further direction — suggested by the study's opening provocation — concerns songs with no semantic content: understanding how listeners construct meaning in the absence of verbal input represents a natural extension of this line of inquiry.

References

Bourne, J. (2025). Who Listens? Experience, Cognition, and Musical Meaning. New York, NY: Oxford Academic.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chilvers, A., Quan, Y., Olsen, K. N., & Thompson, W. F. (2024). The effects of cultural source sensitivity on music appreciation. Psychology of Music, 52(3), 284-304. https://doi.org/10.1177/03057356231201846

Deutsch, D. (1999). Grouping mechanisms in music. In D. Deutsch (Ed.), The psychology of music (2nd ed., pp. 299–348). Academic Press.

Juslin, P. N., & Laukka, P. (2004). Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening. Journal of New Music Research, 33(3), 217–238. https://doi.org/10.1080/0929821042000317813

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. Behavioral and Brain Sciences, 31(5), 559–575. https://doi.org/10.1017/S0140525X08005293

Kiernan, F., Krause, A. E., & Davidson, J. W. (2022). The impact of biographical information about a composer on emotional responses to their music. Musicae Scientiae, 26(3), 558-584. https://doi.org/10.1177/1029864920988883

Morgan, D. L. (1997). Focus groups as qualitative research (2nd ed.). Sage.

Odena, O. (2012). Musical Creativity: Insights from Music Education Research. London: Routledge

Patton, M. Q. (2002). Qualitative research and evaluation methods (3rd ed.). Sage.

Reybrouck, M. (2021). Musical Sense-Making: Enaction, Experience, and Computation. Abingdon, Routledge.

Reybrouck, M., Podlipniak, P., & Welch, D. (2024). Music listening as exploratory behavior: From dispositional reactions to epistemic interactions with the sonic world. Behavioral Sciences, 14(9), 825. https://doi.org/10.3390/bs14090825

Senn, O., Hoesl, F., Jerjen, R., Bechtold, T. A., Kilchenmann, L., Rose, D., & Alessandri, E. (2023). A Stimulus Set of 40 Popular Music Drum Patterns with Perceived Complexity Measures. Music & Science, 6. https://doi.org/10.1177/20592043231202576

Sloboda, J. A., & O'Neill, S. A. (2001). Emotions in everyday listening to music. In P. N. Juslin & J. A. Sloboda (Eds.), Music and emotion: Theory and research (pp. 415–429). Oxford University Press.

Stige, B. (2007). The Grieg Effect – On the Contextualized Effects of Music in Music Therapy. Voices: A World Forum for Music Therapy, 7(3). https://doi.org/10.15845/voices.v7i3.548

Taylor, J.A. (1982). Psychomusicology: Perceptual/Cognitive Research with Implications for the Composer and Performer. College Music Symposium, Vol. 22, No. 1, pp. 90-103. http://www.jstor.org/stable/40374144

Tracey, J. (2011). Benefits and Usefulness of a Personal Motivation Video: A Case Study of a Professional Mountain Bike Racer. Journal of Applied Sport Psychology, 23(3), 308–325. https://doi.org/10.1080/10413200.2011.558364