

THE ANALYTIC-INTERPRETATIVE PARADIGM OF SIGISMUND TODUȚĂ'S LIED SPUNE-O ÎNCET ("SAY IT SOFTLY"): A COMPARATIVE PERSPECTIVE ON HUMAN AND ARTIFICIAL INTELLIGENCE-BASED ANALYSIS WITH ELEMENTS OF COGNITIVE DISSONANCE

Anca BOENGIU¹

¹ Faculty of Music and Theater, West University of Timisoara

¹ anca.boengiu03@e-uvt.ro

ABSTRACT

Human interpretation in music is inherently laden with expressive intent, informed by context, tradition, and individual sensibility. When this expressivity is analyzed by AI, it may lack emotional and expressive understanding and depth in evaluating music, resulting in a form of aesthetic cognitive dissonance (Menninghaus et al., 2017) between the machine's lack of consciousness and its capacity to mimic cognitive functions associated with human mind (Sfetcu, 2021: 6) as well as understanding of all the intricate layers of music interpretation. This paper investigates the interpretive and emotional differences between human and AI analysis of Sigismund Toduță's lieder Spune-o încet, highlighting how cognitive dissonance may be apparent in score analysis when faced with the need for artistic and emotionally-expressive interpretations rendered by non-human agents.

Keywords:

Comparative analysis, artificial intelligence, cognitive dissonance, Sigismund Toduță, vocal repertoire.

INTRODUCTION

The interpretative analysis of a musical work can encompass complex cognitive frameworks, including the tensions between semantic and melodic levels and the broader question of human perception as contrasted with artificial intelligence (AI). Unlike AI, human analysis proceeds through a triadic process rooted in intuition, emotion, and reason, dimensions that together form a conceptual triadic paradigm for artistic interpretation. These elements are operationalized here as:

1. **Intuition** – Defined as “a personal quality for reaching insights or conclusions instantly without conscious reasoning” (Sipman et al., 2019), intuition supports an esoteric and non-discursive approach to interpretation.

2. **Emotion** – Occurring at the affective level (Cloșcă, 2009: 358), emotional perception mediates sensory responses to music and underpins affective engagement with a musical work.
3. **Reason** – Rooted in logical cognition, reason enables analytical dissection through discriminative knowledge and structural awareness.

In contrast, artificial intelligence is confined to the dimension of reason, devoid of intuitive or emotional faculties. This limitation accentuates the value of human-centered analysis, particularly within the ontological dichotomy between biological (analog) and technological (digital) interpretative paradigms.

Sigismund Toduță's lied *Spune-o încet* ("Say it softly") (Toduță, 2006), composed on poetry by Lucian Blaga (Blaga, 1968), transforms vibration – the fundamental property of sound – into music through compositional techniques such as atonalism and rhetorical gesture (Banciu, 2019). Atonalism, characterized by the absence of a tonal center, constructs its logic through pitch-class collections (Alvarez et al., 2010: 82), while rhetorical techniques bridge rhythm and melodic contour. Both approaches ultimately derive from a common origin: the sound itself.

Toduță's subtle harmonic shifts and tonal ambiguities already invite internal interpretive negotiation from performers. When these nuances are mechanically parsed and analyzed by AI, the reader may experience a compounded cognitive dissonance, torn between the musical meaning they assign to affective cues found in the score and the sterile understanding of their meaning (Koelsch, 2014).

Semantic-Melodic Interplay in Human Analysis: Measures 4 and 5

In *Spune-o încet*, the interplay between text and music materializes in semantic and intervallic structures tied to themes of love and existential introspection. The human analytic paradigm reveals that measure 4, through its atonal dissonances in the piano accompaniment, communicates a tension emblematic of the soul's yearning for connection, an inner divergence resulting from the soul's desire for emergent coexistence (measure 4).



The word *Spune* ("Say") is emphasized via crescendo and ascending chromatic motion, while *încet* ("softly") is underscored with a decrescendo and descending semitone, an embodiment of the wordpainting technique.

1234

nu spu - ne ta - re

In measure 5, the semantic variant *nu spune tare* (“don’t say it aloud”) introduces pitch-class [4], augmenting the previous [0123] collection to [01234]. Although semantically opposite, both measures emphasize subtle dynamics, with *tare* (“aloud”) articulated softly to reflect the negation *nu* (“don’t”).

In measure 4, the dynamic *crescendo* is attributed to the word *Spune* (“Say it”), a dynamic development parallel to that of the ascendant melodic chromatic progression denoting, naturally, an increase in intensity, both dynamically and in terms of musical pitches. The dynamic indication *decrescendo* is assigned to the word *încet* (“softly”), an effect that emphasizes a *wordpainting* technique where the melodic progression (descending semitone) is in tune with the text *încet* (“softly”).

By contrast, measure 5 represents the same semantic-musical essence in a different paradigm (textually and with pitch-class augmentation), is destined to be interpreted differently, in *decrescendo*, in the context of the word *nu spune* (“don’t say it”), and the word *tare* (“aloud”) surprises dynamically by using the same intensity as the word *încet* (“softly”). The dynamic interpretation of the word *tare* (“aloud”) transcends the predictable *tare-forte* (“aloud”-*forte*) towards a subtle approach (“softly”-*piano*), a change caused by a change in meaning by association with the negation *nu* (“don’t”). Between the two dynamics in measure 4 (*forte* and *subito piano*) a pause is necessary to maintain the fluidity of the musical discourse:

Spu - ne în - cet nu spu - ne ta - re

This creates a nuanced mirroring between *spune încet* (“say it softly”) and *nu spune tare* (“don’t say it aloud”), distinguished not only by pitch-class extension but also by expressive inflection.

Artificial Intelligence Analysis of Measures 4 and 5

Employing ChatGPT 4.0 for comparative analysis, AI interprets measure 4's pitch-class progression (A–Bb–B–C–B; [0123]) as a gradient of expressive tension followed by resolution: *Spu-* [A, Bb] *ne-o* [B] *în-* [C] *cet* [B] (*Say* [A, Bb] *it* [B] *a-* [C] *loud* [B]), using the primary form (0123). AI analysis emphasizes dynamic shaping, suggesting subtle increases in intensity followed by a soft landing—though it lacks the nuanced semantic alignment present in the human interpretation:

“gradual increases in intensity as the phrase develops, but also a final closure that brings stability. As we move from low intervals (semitones) to a higher interval (pitch), the performer can pay more attention to the expressiveness of the phrase, gradually increasing the dynamics. The final return to [B] can be achieved with a decrease in intensity to mark the closure of the musical phrase:

- The first step (A-Bb): this can be sung with a subtle tension, as the semitone between A and Bb means ultimate closeness, but there is an “uneasiness” that deepens as the phrase continues;
- Second step (Bb-B): this other semitone between Bb and B continues the line of tension, suggesting a desire to 'move forward', but still in a 'zone of uncertainty';
- The third step (B-C): The whole tone between B and C suggests some “opening” or expansion. This can be interpreted as a change of direction, offering a moment of release or increased intensity;
- The last step (C-B): The repetition of the note B suggests stability and closure, as a ‘closure’ of the musical phrase. This can be rendered with a sense of ‘resolution’ and completion of the musical idea” (Chat GPT 4.0).

In measure 5, AI evaluates the [01234] pitch-class set with similar structural focus, emphasizing smooth transitions and the expressive leap at F#-G. However, it overlooks the crucial rhetorical function of textual negation and dynamic inversion – an omission that demonstrates AI’s inability to synthesize semantic nuance with musical structure. The way to interpret a measure within the verse *n-o* [E] *spu-* [F, F#] *ne* [G] *ta-* [G, G#] *re* [G#] (*don't* [E] *say-* [F, F#] *it* [G] *a-* [G, G#] *loud* [G#]), is by using the primary form 01234, as resulting from pitch class analysis, as pointed out by AI:

“the 01234 progression suggests a dynamic that starts with a small step and gradually amplifies, with a larger interval in the middle (whole tone), and then stabilizes with repeated notes:

Semitones (E-F, F-F#) can be played with a smoother and more controlled dynamic, reflecting the gradual approach between notes;

1. The whole-tone leap (F#-G) can be the point of increasing intensity and expressiveness, where the phrase “expands”;

2. Repeating the notes (G-G) suggests a stabilization, a pause in the musical phrase, which may be followed by a slight “build-up” to the G-G# ending:
3. The final closing on G# adds a slight tension that 'ends' on the same note.”

It is important to note the errors made by ChatGPT 4.0 in correctly identifying music intervals and pitch classes (F#-G identified as while tone, G instead of F#, etc.).

Measure 6: Dual Structures in the Human Perspective

Measure 6 exemplifies cognitive dissonance through dual semantic and melodic structures: *lată ațeștia / suntem noi* (“Behold these / we are”). The rhythmic-melodic cell (RMC) is based on two structures that can be distinguished as:

1. Semantic (two words each), determined by the mode of meaning formation *lată ațeștia / suntem noi* (Behold these / we are), and
2. melodic, in terms of the primary form of the pitch classes [A#-B-A] (012) and [F#-E-F] (012) determined once by the semantic level and then by the primary form of the pitch classes.

The dual approach in the RMC analysis represents the cognitive dissonance between the semantic and melodic levels. The segmentation into [0125] and [012] pitch-class sets reflects a diminution from observation to identification.

The image shows a musical staff for Measure 6. The time signature is 5/4, and the dynamic marking is *mp*. The melody starts with a quarter rest, followed by a quarter note (A), a quarter note (B), a quarter note (C), and a quarter note (D). The lyrics are "Ia - tă a-ceș-tia sun tem noi". Red brackets above the notes indicate pitch-class sets: "0125" spans the first four notes (Ia - tă a-ceș-tia) and "012" spans the last three notes (tem noi). There is a caesura after the first four notes.

The secondary compact form (012), represents the diminution of the original primary form (0125). The first segment suggests the idea of a general observation by the composer, *lată ațeștia* (“Behold these”), the cause of the diminution being the brief nominalization of the target characters, *suntem noi* (“we are”).

At the semantic level, the word *lată* (“Behold”) focuses attention on the *status quo* of human potential. Wordpainting is emphasized in *lată* (“Behold”), marked by a tenuto on [A] and ascending semitone, while *ațeștia* (“these”) follows a descending third, evoking disillusionment. The phrase *suntem noi* (“we are”) introduces philosophical duality, highlighting tension between unity and isolation.

The words in the second segment *suntem noi* (“we are”) represent the conclusion of the objective observation of the present situation, in which the idea of duality (*suntem doi* “we are two”) and solitude are outlined from a philosophical point of view (the existence and non-existence of the soul). The interpretation of measure 6 emphasizes verse structure in 3 parts, by means of the caesuras:

6 *mp*

[Ia - tă] [a-ceș-tia] [sun-tem noi]

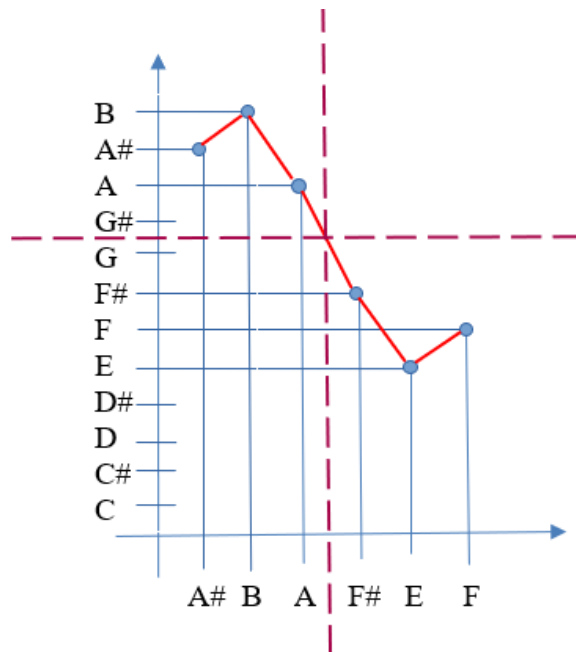
At the melodic level, measure 6 can be structured in two parts: [A#-B-A] (012) and [F#-E-F] (012):

6 *mp*

Ia - tă a-ceș-tia sun-tem noi

At the micro-structural level, the RMC is built from two semitone-tone dyads [B-A] and [A#-B], except for the second structure that is found in the retroversion of the [F#-E] (tone) and [E-F] (semitone) dyads.

The pairs present in the initial form S (semitone) - T (tone) and in the retroversion (tone-semitone), represent an analogy associated with the initials of the composer Sigismund Toduță, a technique that suggests his hidden signature on this meditation on the human condition. The retroversion technique used in the second structure materializes, at the melodic level, the palindrome form, as seen in the analysis of the contour in measure 6, as seen in other pertinent analysis literature (Popean, 2018: 24):



Melodically, the two [012] cells (A#–B–A and F#–E–F) are presented in retroversion and mirror each other, forming a palindromic structure. The dyads within each segment (S–T and T–S) allude, again, to the composer’s initials (S.T.) and represent a musical signature reflective of introspective inquiry.

AI Analysis of Measure 6

ChatGPT's response segments measure 6 into two expressive blocks, correctly identifying semitonal tension followed by stabilization interpreting the verse *la- [A#] tã [B] {a- [A] ceș- [A] tia} [F#] {sun- [F#] tem [E, F]} noi [F] (Be- [A#] hold [B] {these [A - F#]} are {[F#-E-F]} we [F])*, using the primary forms [(012) and (012)] resulting from pitch-class analysis:

“the interpretation of this musical measure involves an approach to the progression in two distinct blocks, each having a gradual movement between intervals. The first sequence has a smooth, semitone progression that creates a subtle tension, while the second sequence introduces larger leaps and then establishes the phrase by repeating notes. Thus the performance should emphasize the subtle movements between notes, but also the expansion and stabilization of the phrase through the larger leaps.

- The semitones (A# - B, B - A) can be played with a smooth and delicate movement, being slightly dynamic to build tension progressively;
- The whole-tone leaps (A - G, G - E) can be played with a more pronounced dynamic, signaling an expansion of the phrase;
- The repetition of notes (A - A, G - G, F - F) provides moments of balance and stabilization in the phrase, which can be emphasized by a decrease in intensity” (ChatGPT 4.0).

It is important to note the ChatGPT 4.0 errors in identifying intervals and tones correctly (B-A as a semitone, G-E as a whole tone, G instead of F# in several instances) that makes such an analysis prone to wrong conclusions.

However, its interpretation lacks philosophical insight and fails to address Toduță’s rhetorical devices or the intertextual resonance with Blaga’s poetry. AI perceives structural balance but not its symbolic or affective implications.

Artistic interpretation through atonalism synthesizes reason and intuition, while rhetorical analysis links reason with emotion, bridging the esoteric and the analytical. Affect theory, central to rhetorical strategy, underscores the potential of sound analogies (e.g., rhetorical figures) as vehicles for complex emotional expression, from joy to sadness, from euphoria and optimism to melancholy (sensory-based music analysis) (Arbore, 1984).

Measure 12: The Fragility of Duality – A Human View

The line *Stăm alături eu și tu?* (“Are we standing side by side, you and I?”) encapsulates existential uncertainty in duality through rhetorical figures such as the *rhetorical question* and the *saltus duriusculus*. The rhetorical question (*interrogatio*) invokes introspection rather than inquiry, heightened by pronoun inversion *eu și tu* (“me and you”) instead of *tu și eu* (“you and I”), a signal of cognitive imbalance within the dyadic relationship.

The rhetorical question *Stăm alături eu și tu?* (“Are we standing side by side, you and I?”) is semantically surprising due to the inversion of the pronouns *eu / tu* (“me” / “you”), a process that transcends the original nature of their normal use *tu / eu* (“you” / “me”). The inversion highlights the secondary character *tu* (“you”) as the main factor of the imbalance in duality, due to the illusion inherent in the subject-object cognitive paradigm. The idea of fragility is captured at the melodic level in the inversion technique of *eu și tu* (“me and you”): the distance between the two sounds is not a semitone that, in the initial state, would support the concept of *alături* (“beside”, meaning “side by side”), but is whole tone *eu-B / tu-C#* (*me-B / you -C#*), creating contrast between the music and the verse. The B (*eu* “me”) and C# (*tu* “you”) create rhythmic instability: B is on the upbeat but in syncopated form while C# occurs on the downbeat.

1 2 3 4 5 6

12

Stăm a-lă-turi eu și tu?

Saltus duriusculus is a dissonant interval (Bartel, 1997: 381), which in this context is a descending interval of the augmented C#-G. The *wordpainting* technique is presented by associating the concept of *alături* (“side by side” of “beside each other”) emerging from the Romanian word *alături* (“beside”) with the rhetorical figure *saltus duriusculus* that emphasizes the existence of an unstable dual relationship suggested by the dissonant interval of augmented fourth. The existential situation is underpinned by uncertainty, shaped by the rhetorical question. *Saltus duriusculus* is preceded by pitch class D that, in retroversion, represents the resolution of the augmented fourth-induced dissonance. The word *alături* (“side by side”) represents the center of interest in the measure, being presented in the regression-enhanced intervallic scheme: D-C#-G (2m-4+).

12

4+

Stăm a - lă - turi eu și tu?

Measure 12 can be segmented into several analytic layers, such as interpretative and melodic, resulting in 3 segments each, delineated by caesura, each time based on a different rationale

From a melodic standpoint, the 3 segments are represented by 3 dyads, each based on different intervallic motion starting from pitch class C#: C#-D, C#-G, and B-C#. Musically, the C#-G interval (augmented fourth, or *saltus duriusculus*) dramatizes this fragility, preceded by D in retrograde motion to resolve dissonance. The phrase centers on *alături* (“side by side”), segmented into dyads reflecting C#'s role as a constant, with D, G, and B forming a variable contour signifying instability.

12

Stăm a - lă - turi eu și tu?

From an interpretative point of view, the 3 segments take into account the placement of caesura based on semantic rationale as this approach segments the melodic level as applied to words:

12

Stăm a - lă - turi eu și tu?

Melodic segmentation into C#-D, C#-G, and B-C# reveals dual tensions across each unit, illustrating unresolved dissonance between semantic and musical structures. The C# pitch acts as an anchor (representing “I”), while other pitches suggest emotional or relational divergence (“you”).

An exception to the first dyad is made within the melodic caesura: while melodically it must be placed after the first dyad {*Stăm a* - [C#-D]} (“Are we standing side by...”), here it appears into the middle of the first dyad, between the two pitches.

Again, each segment showcases a different interval based on of pitch class C#: C#-D (2m), C#-G (4+) and B-C# (2M). The latent divergence between the semantic and musical layers creates novel texture by superimposing the latent inner rhythms of each layer.

Duality as a principle is conveyed in each segment by the constant presence of the C# sound, forming pairs of sounds [C#-D (2m), C#-G (4+) and B-C# (2M)]. The relationships' dual nature fragility and uncertainty is musically conveyed by the unresolved tensions of dissonant intervals.

Pitch class C# represents a melodic constant, an element of certainty associated with the word *eu* "I", while D, G and B suggest a rhetorical counterpoint and insecurity attributed to the external agent *tu* "you" by variable melodic contour, contributing to the innate sense of imbalance:

12

Stăm a - lă - turi eu și tu?

One notable exception is found in the third segment where the function of pitch classes B and C# attributed to the words *eu* "I" and *tu* "you" are inverted: the B sound is attributed to the word "I" while the C# sound is associated with the word "you". The interchange of the role between the two pitch classes is emphasized by the rhetorical question, not denoting a categorical statement but a hypothesis.

Measure 12: Artificial Intelligence Interpretation

ChatGPT identifies rhetorical and dissonant devices, correctly naming *interrogatio* and *saltus duriusculus*, and recognizes G-B as a dissonant leap. Nonetheless, it fails to integrate this into a broader interpretative framework. The inversion of roles between *eu* "I/me" and *tu* "you", and the instability created by rhythmic positioning (syncopation vs. downbeat), are not acknowledged, highlighting the limitations of AI in capturing ontological and poetic depth. Analysis of measure 12 was based on the *Stăm* [C#] *a-* [D] *lă-* [C#] *turi* [G] *eu* [B] *și* [B, C#] *tu?* [C#] (*Are we standing* [C#] *side-* [D] *by-* [C#] *side* [G] *me* [B] *and* [B, C#] *you?* [C#]) verse, using the rhetorical figures *interrogatio* and *saltus duriusculus*:

"the rhetorical figure *interrogatio* is used when a rhetorical question is asked that does not expect an answer, but is meant to emphasize an idea or draw attention to a theme. In the case of *Stăm alături eu și tu?* the question *tu?* is phrased in a way that suggests a dialog or relationship between speaker and addressee, but without implying a real demand for an answer. The question in this context may reflect a sense of uncertainty or expectation, a kind of withdrawal in the face of a decision that is not clearly stated explicitly. This question can be perceived as a challenge or an invitation to reflection, intended to provoke thoughts about the nature of the relationship between the subject and the speaker. The figure *saltus duriusculus* is used to describe a sudden and

difficult leap between two notes, usually with a fairly long interval between them. In the case of the lines *Stăm (C#) a- (D) lă- (C#) turi (G) eu (B) and (B, C#) tu? (C#)*, there is a rapid change between different pitches and a jump between notes that can create an effect of tension. This jumping between notes can be seen as a technique to emphasize emotional contrast or conflict within the text. For example, the leap from G to B could create a pitch difference that emphasizes the dilemma or uncertainty of the final question *tu?*".

In terms of the proposed analytical model, artistic interpretation is considered as having both essence and contour. Both perspectives (both human and artificial) represent the same contour constructed by reasoning. The artificial intelligence, although it only possesses the possibility of logical thinking, does not present correctness in terms of the quality of information (mistakes both at the semantic level (spelling and expression) and at the musical level through erroneous statements). In contrast, the essence of analysis which is constituted by vibration and emotion underlies the phenomena of cognitive nature accessible to the human agent but inaccessible to the artificial paradigm, an effect caused by the impossibility of intersection of the artificial with intuition and emotion.

Conclusion

Within the proposed analytical model, artistic interpretation encompasses both essence and contour. While both the human and AI perspectives are capable of outlining structural contours via rational processes, only the human approach accesses the interpretative essence, grounded in intuition and emotion. Artificial intelligence, bound by its logical framework, cannot penetrate the affective and intuitive dimensions fundamental to the human experience of music.

This contrast underscores the irreplaceable nature of human analysis in understanding and conveying the multifaceted meaning of musical works, particularly when embedded in philosophical and poetic contexts such as those presented in Toduță's lieder *Spune-o încet*.

REFERENCES:

- Alvarez, Inma, Hector J. Perez, Francisca Perez-Carreno. 2010. Expression in the performing arts. UK, Newcastle upon Tyne: Cambridge Scholars Publishing.
- Arbore, Grigore. *Forma ca viziune. Considerații asupra barocului*. Editura Meridiane, 2011.
- Banciu, Ecaterina și Gabriel Banciu. 2019. *Retorica muzicală*. Curs. Biblioteca Academiei Naționale de Muzică „Gheorghe Dima”.
- Blaga, Lucian. (1968). *Mirabila sămânță*. Editura pentru literatură.
- Cloșcă, Ciprian Marius. (2002). *Docta Ignorantia*. Editura Lumen.
- Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience*, 15(3), 170–180.
- Menninghaus, W., Wagner, V., Wassiliwizky, E., Schindler, I., Hanich, J., Jacobsen, T., & Koelsch, S. (2017). The distancing-embracing model of the enjoyment of negative emotions in art reception. *Behavioral and Brain Sciences*, 40, e347.
- Popean, Mihai. *20th-Century Music Analysis Case Studies*. Cluj-Napoca, Editura Media Musica, 2018.
- Sipman, Gerbert, Jürg Thölke, Rob Martens, Susan McKenney. (2019). The role of intuition in pedagogical tact: Educator views. BERJ - British Educational Research Journal. BERA Journals. Accessed May 5, 2025, <https://doi.org/10.1002/berj.3557>.
- Sfetcu, Nicolae. *Introducecere în inteligența artificială*. MultiMedia Publishing, 2021.
- Wink, A. Karen. *Rhetorical Strategies for Composition. Cracking an Academic Code*. Rowman & Littlefield Publishers, 2020.
- Toduță, Sigismund. (2006). 5 lieduri pentru voce (mezzosoprană) și pian 1983-1988. Editura Arpeggione.