Quantifying Musical Micro-Motion: Applying Optical Flow Analysis to Audiovisual Abstract Animation

Gerald Moshammer
Humanities and Language Division,
Mahidol University International College,
Thailand
gerald.mos@mahidol.ac.th

Abstract

Optical flow analysis, while initially developed for computer vision, has expanded its applications into various domains. While traditional cognitive models of musical expression focus on understanding the impact of music on listeners, they often show less interest in what the music itself expresses or fail to fully elucidate the complexities associated with this sphere. Moreover, these models tend to be coarse and predominantly top-down, categorising music based on standard emotion theories using general musical dimensions, without considering the intricacies of perceptual abstraction and idiosyncratic processing. In contrast, contemporary computer software allows for a more nuanced exploration of music aesthetics by integrating music visualisation into optical flow analysis. This article explores the potential of this integration, extending the application of algorithmic music visualisation to serve as a tool for feature detection and audiovisual priming at a finer scale. Optical flow analysis facilitates the collection and evaluation of data that transcends intuitive audiovisual experience, offering a method which may lead to a more "objective" understanding of aesthetic dimensions. By analysing micro-motion and micro-expression in abstract animations of the acoustic spectrum, this approach opens avenues to subtleties previously unexplored and deemed "ineffable". Through two case studies involving Aphex Twin's "Bucephalus Bouncing Ball" and Frédéric Chopin's Prelude op. 28 no. 3, this article demonstrates the potential of optical flow analysis in uncovering hidden layers of meaning in musical expression via idiosyncratic, animated representations of spectral data.

Keywords: Optical flow, abstract animation, musical motion, musical expression, audiovisual priming.

1 Introduction

In psychoacoustics, the concept of the Just Noticeable Difference (JND), or alternatively, the differential threshold, denotes the smallest discernible change in a stimulus that a listener can perceive (see Gescheider, 2013). Linguistics delineates this specific boundary between perceptibility and imperceptibility through the construct of a phoneme—an elemental sound unit in a language capable of altering meaning. In the field of musicology, terms such as "museme" (Tagg, 2012) or "meme" (Jan, 2007; Gofton, 2017), have been introduced to extrapolate phonetic principles into the domain of music. Here, it is noteworthy that perceptual potentiality engages in a dynamic interplay with acoustic or syntactic abstraction, not only in relation to a particular musical language or symbol system but notably in the development of a specific genre or style. Subsequently, aesthetic attention involves the creative or deliberate selection, configuration, and perception of specific syntactic elements in the artistic or creative process, to elicit particular modes of symbolisation.

Ludwig Wittgenstein (1966) may be considered the *locus classicus* for philosophical scrutiny regarding these matters, delving into compelling inquiries concerning the interplay of aesthetic appreciation, symbolic differentiation, and the nuanced selection of “material” in artistic practice. More broadly, perceptual learning has become a focal point of extensive study, situated at the intersection of attention, discrimination, and judgement. Recent research, for instance, has shown that expert musicians often demonstrate enhanced perceptual learning abilities compared to non-musicians, particularly in tasks related to pitch, timing, timbre, and musical structure (see Margulis, 2013; Peretz et al., 2015; Wong et al., 2007; Zendel & Alain, 2012), which also may influence the processing of emotions (Juslin & Västfjäll, 2008; Koelsch, 2014). It is worth noting, however, that some studies have suggested a more critical perspective, such as Mosing et al. (2014), who found no causal effect of music practice on music ability, yet emphasise heritability instead. Despite such findings, expert musicians often seem to exhibit superior attentional control and selective attention skills compared to non-musicians, allowing them to focus more effectively on musical details, as well as demonstrate superior acoustic discrimination skills compared to non-musicians, particularly in tasks involving fine-grained auditory processing (Habibi et al., 2016; Kraus et al., 2014; Strait & Kraus, 2011). In Alexander Truslit’s seminal work on musical motion, the diverse building parameters of music are succinctly underscored. Truslit asserts that “the acoustic elements that an artist manipulates in shaping a performance are pitch, timbre, intensity, and duration; while the first two are of great importance for the composer, the last two are most important for the performer” (Repp, 1992, p. 267). Consequently, music often exhibits a “double life”, existing as discrete entities in notation while manifesting densely idiosyncratic qualities during performance.

In his analytical taxonomy of symbol systems, Nelson Goodman (1976) employs the logical framework of notational systems to delineate syntactic and semantic attributes that facilitate a sharp differentiation between score-based music and verbal or pictorial symbol systems. Goodman’s rationale underscores the unambiguous definition of an equivalence class of correct performances by music scores, hence underlining the structural layer of music’s symbolism. Indeed, given that the score itself resides within such a class defined by pure discrete structural information, one can theoretically reconstruct an authoritative music score based on a “correct” musical performance coupled with the application of the appropriate notational system.

Goodman’s logically idealistic characterisation of music operates primarily within the domain of what he terms a “two-stage art” (Goodman, 1976). However, even within the ostensibly limited realm of score-based music, one must not overlook the historical and aesthetic intricacies inherent in the dynamic interplay between score information and performance practice (Cook, 2013). The defining role of musical notation and its realisation through performance gains further complexity when scrutinised through the analytical lenses of classical music theory in juxtaposition with, say, ethnomusicology, or views in postcolonial music analysis (Nettl, 2005; Titon et al., 2008).

Nonetheless, it is imperative to acknowledge the irrefutable reality that in Western music tradition, notations of musical sound in, for instance, an ear training test typically admit only one correct answer—a claim not asserted when a student is tasked with meaningfully describing or depicting an object. Furthermore, the discrete structural notational foundation inherent in specific musical genres plays a pivotal role in shaping the aesthetic conception of music and the construction of compositions. This structural foundation significantly influences the manner in which individuals engage with and analyse musical works, which, however, cannot take away from the fact that at the end, musical sound, in its essence, unfolds within what Goodman terms a “dense system” that is continuously valued (Goodman, 1976).

This duality highlights the richness of musical sound as one of the most intricate information systems engaged with by human beings. A single tone possesses the capacity to incorporate four separable parameters simultaneously (pitch, dynamics, timbre and duration). Moreover, the combination of musical sounds results in the generation of overlapping higher-level properties, such as melody, harmony, rhythm, and texture (see Temperley, 2019). The remarkable ability of the human mind to synthesise such informational complexity and diversity into a coherent stream of meaningful musical sequences is itself noteworthy (see Levitin, 2006). As a temporal art form, however, music also relies on the potential acuteness of human listening to detect sonic changes and differentiations. The human auditory system has a high temporal resolution, allowing it to distinguish between short intervals and accurately process the timing of auditory events (see Poeppel, 2003; Grondin, 2010). Precision and definiteness of tonal and rhythmic elements, coupled with the immediacy and inescapable presence of aural stimuli, contribute essentially to music's unique cultural and aesthetic function.

Classical music enthusiasts, for instance, exhibit a notable proclivity not only for the aesthetic appreciation of musical compositions but also for an analytical exploration of performances, revealing an intricate engagement with the art form. This phenomenon is particularly pronounced in the meticulous scrutiny of seemingly minor distinctions in the execution of musical works, a tendency observable in the discerning world of music competitions. Within this milieu, judges, often esteemed musicians, endeavour to discern optimal renditions based on nuanced parameters, transcending mere technical proficiency to unveil interpretative subtleties. As of January 2024, the online platform *Wikipedia* enumerates 107 international piano competitions, predominantly catering to young virtuosos, including, for instance, the prominent International Chopin Piano Competition, International Tchaikovsky Competition, Queen Elisabeth Music Competition and Van Cliburn International Piano Competition.[1] This not only underscores the competitive nature of the classical music domain but also illuminates the substantial investment of time and effort by numerous individuals in both the creation and evaluation of the intricate expressive and stylistic facets within the realm of music at an unparalleled level of expertise and differentiation.

The paramount role of sound takes precedence in musical genres devoid of conventional notation, as observed in, say, electronic music, which lacks immediate ties to notational structures. This departure from score-based conventions finds a notable exemplar already in "musique concrète" (Xenakis, 1992; Schaeffer, 2012). More broadly, it is probably not an exaggeration to state that contemporary amateur listeners will likely consider neither music notation nor more general music theoretical concepts in any depth during their appreciation of popular music (see DeNora, 2016; Kahn, 1999; Cox & Warner, 2004).

Adopting a philosophical perspective, the inherent informational richness of music prompts contemplation on mental projections encompassing abstraction, attention, and gestalt creation derived from auditory stimuli. This psychological phenomenon aligns with principles derived from gestalt psychology, wherein perceptual experiences are seen as holistic configurations rather than mere aggregations of individual elements (see Wertheimer, 1923; Köhler, 1947). Moreover, the predictive coding paradigm in cognitive science provides an insightful framework, suggesting that the brain actively generates predictions about sensory input, contributing to the construction of a coherent perceptual unit (Friston, 2005; Clark, 2013). Within this framework, the continual negotiation between identity and difference in cognition and perception is viewed as an ongoing process of refining predictions and updating mental models (Hohwy, 2013; Feldman & Friston, 2010). In the context of music and its perception, this

---

[1] List of classical music competitions. (n.d.). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/List_of_classical_music_competitions.

negotiation prompts an essential question: to what extent do specific alterations in musical parameters impact the expressive facets, semantic dimensions, and communicative functions of the musical experience? This inquiry serves as a crucial avenue for comprehending the intricate connections between the structural elements of music, the perceptual responses they evoke, and the cognitive processes that underpin musical interpretation.

Although there are comprehensive and systematic models addressing music's affective impact and expressive qualities, these are characterised by two limiting features. First, such models appear mainly to be situated within "reception aesthetics", i.e., they are merely interested in the impact music has on its listeners (Gabrielsson & Lindström, 2010). Hence, these frameworks deal with what the listener feels and not necessarily with what the music expresses. Second, such models are coarse and, in terms of categorisation, "top down", particularly in employing terminology from standard emotion theories (Juslin, 2013). Alternatively, more concrete performance studies often operate purely bottom-up in their analysis of acoustic data, without, however, substantially acknowledging the role of perceptual abstraction, mental projection, and idiosyncratic top-down gestalt processing (see Cancino-Chacón et al., 2018, for data-driven approaches). To put it bluntly, in terms of contributions to genuine music aesthetics, the science of music appears to seek to bolster its foundational strength to large degrees either in a rather rigorous attachment to empirical positivism that meticulously accounts for acoustic data, or else imports theoretical frameworks from general emotion theories (see, paradigmatically, Scherer, 2004).

Contemporary computer software, however, opens avenues not only for the retrieval, evaluation and categorisation of sets of big data, it also can assist in gaining valuable aesthetic insights into what is structurally happening at the expressive level that may underpin the appreciation of music. Particularly a temporal art form like music where possible multi-layered sound comes and goes, is hard to pin down from the perspective of what actually constitutes its perceptual and cognitive experience. Here, modern technologies catalyse fresh explorations. In this vein, this article reinterprets concepts from Moshammer (2023), integrating music visualisation into optical flow analysis. This approach extends the application of algorithmic music visualisation beyond its conventional role, serving as a tool for feature detection and audiovisual priming at a finer scale. Recognising the heightened sensitivity of hearing to aesthetically relevant aspects, particularly in rhythmic and motion perception (see Repp & Penel, 2004; Iordanescu et al., 2013), which is more specifically captured by the so-called auditory-motor enhancement hypothesis (Ammirante et al., 2016), optical flow analysis facilitates data collection and evaluation that transcend mere intuitive audiovisual experience. It thus offers a pathway to an "objective" understanding of aesthetic dimensions that would otherwise be accessible primarily through the act of competent or idiosyncratic listening.[2] Unlike approaches centred on decoding the basic bodily motions of performers, listeners, or dancers, this method is grounded in the direct abstract animation of the acoustic spectrum. By employing optical flow analysis at an unprecedented level of detail, it delves into the analysis of micro-motion and micro-expression, illuminating subtleties previously unexplored.

Optical flow analysis, a cornerstone of computer vision, has widened its disciplinary boundaries. It now finds applications in diverse fields such as dance analysis (Luo & Ning, 2022), music performance evaluation (Visi, 2017), autonomous vehicle navigation (Guizilini et al.,

---

[2] The term "objective" here is used in a limited, methodological sense. It refers to the possibility of intersubjective accessibility: animated representations based on the geometrisation of musical spectra can reveal fine-grained motion and dynamic detail that are typically available only to expert listeners through perceptual learning and cultivated aesthetic competence. By externalising these micro-temporal features visually, such methods enable a shared analytical and experiential ground that bridges individual expertise and broader interpretive accessibility.

2022), and even medical research (Liu et al., 2019). For instance, in dance analysis, researchers have utilised optical flow analysis for automatically analysing human movement in dance videos (Luo & Ning, 2022). In the domain of music performance evaluation, optical flow has been explored for automatically assessing music performance quality (Visi, 2017). While its initial theoretical development, as documented by Niehorster (2021), might seem far removed from the realm of music aesthetics, a closer look reveals a profound opportunity. Optical flow algorithms estimate the motion of pixels between consecutive frames in a video sequence, typically facilitating object tracking or motion pattern analysis. In the context of animated music, however, the "objects" under analysis are not physical entities but rather visual representations of sound.

Particularly the confluence of algorithmic music visualisation and optical flow analysis invites a philosophical exploration of how music expresses itself. This goes beyond mere multisensory fusion, suggesting convergence of sound and visual motion as fundamental elements that co-create a unique platform for expressive transformation. Within this framework, music emerges as an intrinsic medium capable of articulating entirely new forms of (e)motionality, ushering in a conceptual space where music becomes a dynamic entity shaping its own expressive trajectory, free from the constraints of physical movement on the part of the performer. It is important to acknowledge that both algorithmic visualisation of the musical spectrum and optical flow analysis are quantitative methods that allow for the exploration of aspects of music previously considered ineffable or beyond measurement. Applying optical flow analysis to automated abstract music visualisation delves into the nuanced intricacies of musical expression, potentially revealing hidden layers of meaning that traditional interpretations might miss.

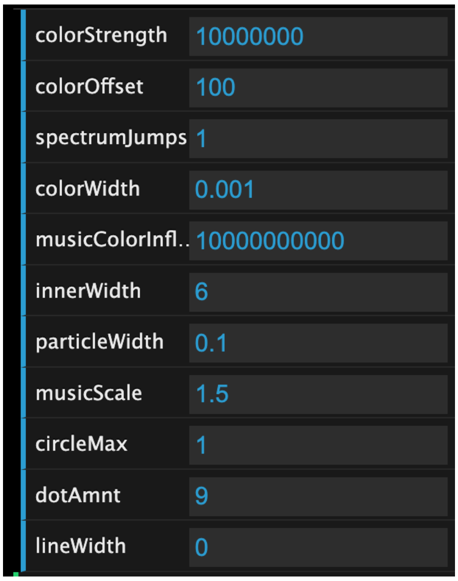
In the remaining sections of this article, I discuss and contextualise the Google Chrome extension *AudioVisualizer*, as introduced, applied, and explained in Moshammer (2023), in conjunction with *FlowAnalyzer*, an optical flow detection tool provided by CEFALA (Centro de Estudos da Fala, Acústica, Linguagem e músicA) detailed on the centre's webpage.[3] I reference Moshammer (2023) for specifics regarding *AudioVisualizer*, and emphasise key aspects of optical flow analysis as necessary throughout my analysis. Furthermore, two case studies are presented to evaluate the potential and constraints of the suggested methods. Firstly, an excerpt from Aphex Twin's "Bucephalus Bouncing Ball", an iconic piece of electronic music also featured in the 1998 conceptual psychological thriller "Pi", is utilised to illustrate the subtle aesthetic intricacies present, even within what may seem like a series of simple bass drops or musical punches. Secondly, I conduct a comparative analysis of three performances of Frédéric Chopin's Prelude op. 28 no. 3, extending the analysis into the domain of performance studies. This comparative analysis focuses on both overall textural features and pinpoints the meticulous evaluation of detailed performance gestures. Lastly, a brief evaluative section reflects on the notions of trans-modal and algorithmic hermeneutics to conclude the article.

---

[3] *FlowAnalyzer* was previously available at https://www.cefala.org/FlowAnalyzer/, though the page appears to be no longer accessible. Researchers seeking to replicate this workflow may contact the original developers or use equivalent optical flow tools. The software has also been employed in other contexts, such as Hall et al. (2017).

2 A Moving “Bucephalus Bouncing Ball”: Restitution, vertical Acceleration and Expression

Aphex Twin’s “Bucephalus Bouncing Ball”[4] has a duration of approximately 5 minutes and 44 seconds. Between approximately 1 minute and 33 seconds and 1 minute and 54 seconds, the piece includes a section comprising eight groups of four bass drops, which will be analysed in further detail. To introduce the passage and aid an intuitive understanding of how audiovisual flow analysis manifests, consider Animation 1, based on the *AudioVisualizer* animation profile as depicted in Fig. 1. This animation superimposes a data trail illustrating the magnitude of motion in terms of optical flow within the specified animation region (see Fig. 2), generated by *FlowAnalyzer*.



colorStrength	10000000
colorOffset	100
spectrumJumps	1
colorWidth	0.001
musicColorInfl..	10000000000
innerWidth	6
particleWidth	0.1
musicScale	1.5
circleMax	1
dotAmnt	9
lineWidth	0

Figure 1: Settings in *AudioVisualizer*’s “DotsAndLines” profile as used in Animation 1.

Optical flow analysis enables focusing on a specific region of the screen, allowing for the examination of motion flow within that selected area only. The data are presented in reverse order to align the accelerating downbeats with increasing velocity and were exponentially smoothed in *Excel* using a factor of 0.9. All subsequent exponential smoothing featured in this article was also performed in *Excel*. Additionally, “reversed data” refers to pixel displacement on the x-axis, where higher quantities are shown in lower positions and *vice versa*, to align more intuitively with the trajectories and velocity patterns of the moving images.

One immediately notices the substantial amount of distinctive data revealed through optical flow analysis, given the common rate of 60 video frames per second. At first glance, this is promising, as it demonstrates that optical flow analysis of abstract music animations, through the geometrisation of the musical spectrum, provides insights and features that acoustic analysis alone cannot directly uncover. However, it is important to note that optical flow analysis measures pixel changes from one frame to the next, in Cartesian or polar coordinates. Therefore, not every visually detected change can be straightforwardly assigned musical significance when evaluating musical motion. To enhance the readability of data graphs and to create trendlines that are both meaningful and sufficiently detailed to reflect aural nuances, applying degrees of exponential data smoothing is beneficial throughout the analysis.

---

[4] First published on: Aphex Twin (1997). *Come To Daddy*. Warp Records (WAP94CDX).

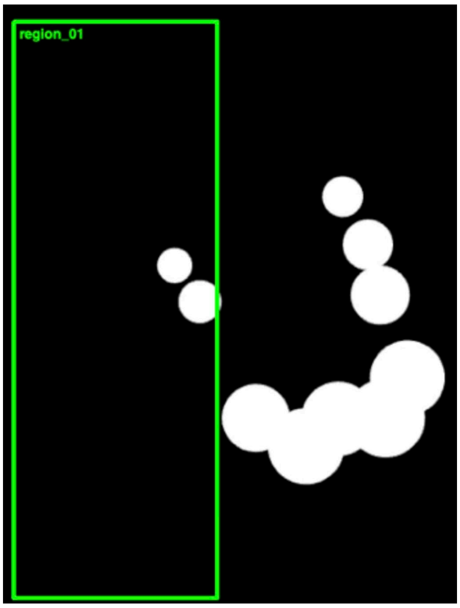


Figure 2: Animation region used in *FlowAnalyzer*'s analysis of motion magnitude that underpins Animation 1.

Animation 1: Visualisation of the "Bass Beat Section" from Aphex Twin's "Bucephalus Bouncing Ball" with *AudioVisualizer*'s "DotsAndLines" profile (see text for further details).

One approach to creating a more immediate graphical representation of flow data involves depicting cumulative vertical or horizontal pixel displacements. This method aims to generate a contour of movement that is spatially more intuitive in both direction and magnitude compared to the raw data from optical flow analysis in terms of overall motion magnitude. Interestingly, in the cases discussed in this article, this display of cumulative data reveals a fundamental slope in the overall data trails that should not theoretically appear if the animation only fluctuates without changing its centred positioning. However, when reading vertical and horizontal pixel replacements in changing areas, such as the expansion of a circle area that moves downwards (and vice versa), optical flow analysis tends to assign asymmetric quantities to upward and downward motion. These asymmetries are rooted in the particular algorithm and mapping that underpins the automated abstract animation of musical spectrum. This results in a specific yet intrinsically consistent sloping of the data graphs in optical flow analysis that uses cumulative cartesian coordinate data, which can be viewed as evidence for the method's coherent visual evaluation of a particular audiovisual animation profile (further discussion can be found in the comments regarding Fig. 14 below).

Another crucial aspect of interpreting optical flow data is to emphasise the relative values of motion magnitude and pixel displacement, rather than placing excessive emphasis on absolute numbers. This focus on relative change is essential, especially when considering abstract animation where the exact algorithmic details of flow calculation are less critical compared to tracking real objects. In abstract animation, the objective is to capture the dynamism behind audiovisual images. One of the main reasons for prioritising the evaluation of relative change is the ability to adjust the scaling of the analysis simply by enlarging or minimising the scanned footage.

Fig. 3 below presents a comparison between the overall motion data of the region depicted in Fig. 2 using polar coordinates, which also informs Animation 1, and, additionally, accumulative vertical motion data. The latter distinctly delineates the groups of four beats, as evident from the analysis of a small visual portion of the animation (refer back to Fig. 2) that specifically references the base beats. This clarity and uniformity are further emphasised by the zoomed-in comparison depicted in Fig. 3, distinguishing between what is referred to as Group A and Group E. These labels correspond to the score transcription of the analysed excerpt from "Bucephalus Bouncing Ball" (see Fig. 4), which outlines eight structural units of the composition defined by the groups of four-beat drops in the passage. Group E plays an idiosyncratic role in this passage due to its unique rhythmic structure, which will be highlighted in the subsequent analysis.

At this point, one might question whether optical flow analysis provides intrinsic benefits over simply listening to music or conducting acoustic analysis. In the straightforward case of a series of bass drops, the clear motion pattern might not seem to justify a complex analytical approach. However, this simple case effectively illustrates the value of visualising the acoustic spectrum and highlights how animated visuals can enhance our understanding of musical motion. The animated result may indeed be connected to music listening, but both the listening experience and visual animation contribute additional insights beyond the intrinsic acoustic features alone. For instance, the circle that appears to be thrown and accelerated downward can be likened to the motion of a basketball, demonstrating how visual representation can evoke associations that extend beyond the acoustic properties of the music. This contributes to our understanding of how the human mind interprets musical dynamics.

More fundamentally, the animation of sound – specifically the visual representation of sound's dissipation after a bass drop – is shaped by the chosen algorithmic method for geometrising the musical spectrum, as implemented by *AudioVisualizer*. This method translates acoustic data into visual form, creating a continuous trajectory that may be seen as virtually mimicking performance gestures and highlights the preparation for subsequent bass accents. This effect is particularly noticeable during the extended intervals between bass drop groups, where the circles representing the strikes move further upward, suggesting a more energised subsequent drop. While this animation method effectively mirrors the dynamic outline of the beat groups, it also introduces a keyframe to the motion trajectory, visually depicting the ball being held in an elevated position before being thrown downward. This example demonstrates how abstract music animations, by spatially representing musical spaces, produce emergent motion and expression that are not immediately deducible from the acoustic image alone. While the visual interpretation may seem natural, it is rooted in the mental experience of music listening and is enhanced by the visual representation provided by the algorithmic method.

The interplay between abstraction and the property of emergence is crucial in this context. In addition to spectrum jumps, by adjusting the number of dots that underpin the animated shape's outline, even within the relatively straightforward algorithmic framework employed by *AudioVisualizer*, one can achieve a variety of diverse yet structurally interrelated visual gestural profiles corresponding to an identical sound image. This variability in dot quantity and placement enables a rich array of visual representations, each uniquely capturing the essence of the musical experience from a dynamic and energetic viewpoint (see Animations 12 and 13 in Moshammer, 2023, p. 20, for exemplification).

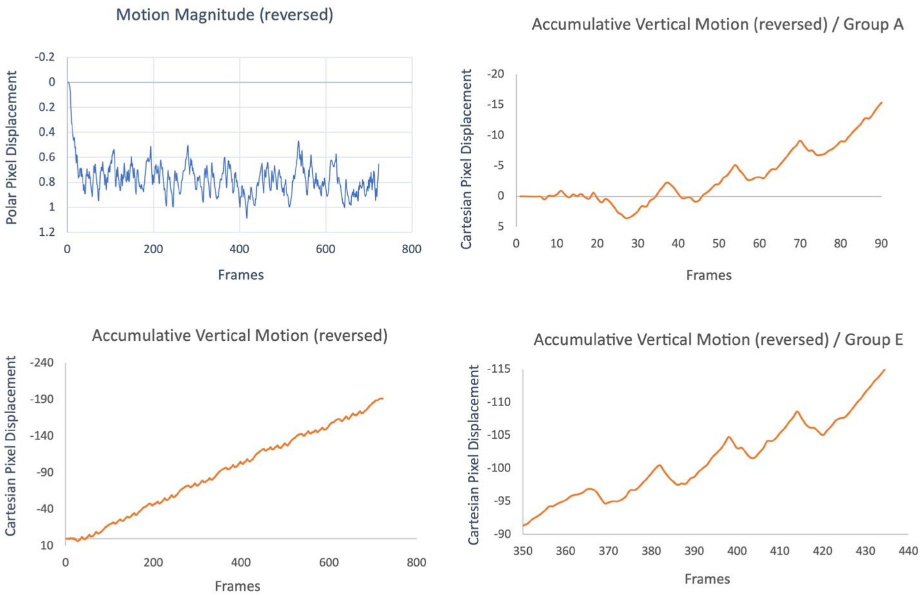


Figure 3: Comparison of the animation region motion magnitude data as shown in Animation 1 with the respective vertical motion data in *FlowAnalyzer*, equally smoothened with a factor of 0.9, with accumulative vertical motion data (see text for details).

Before delving into the intricacies of beat section E's motion, however, I would like to maintain a bird's eye view of the whole passage under discussion by directing the reader's attention to Animation 2. This animation compares the original version of the discussed passage with a MIDI rendering of its approximate transcription displayed as Fig. 4, by mapping these two acoustic versions onto their matching and mismatching animations in *AudioVisualizer*, again based on the profile shown in Fig. 1.

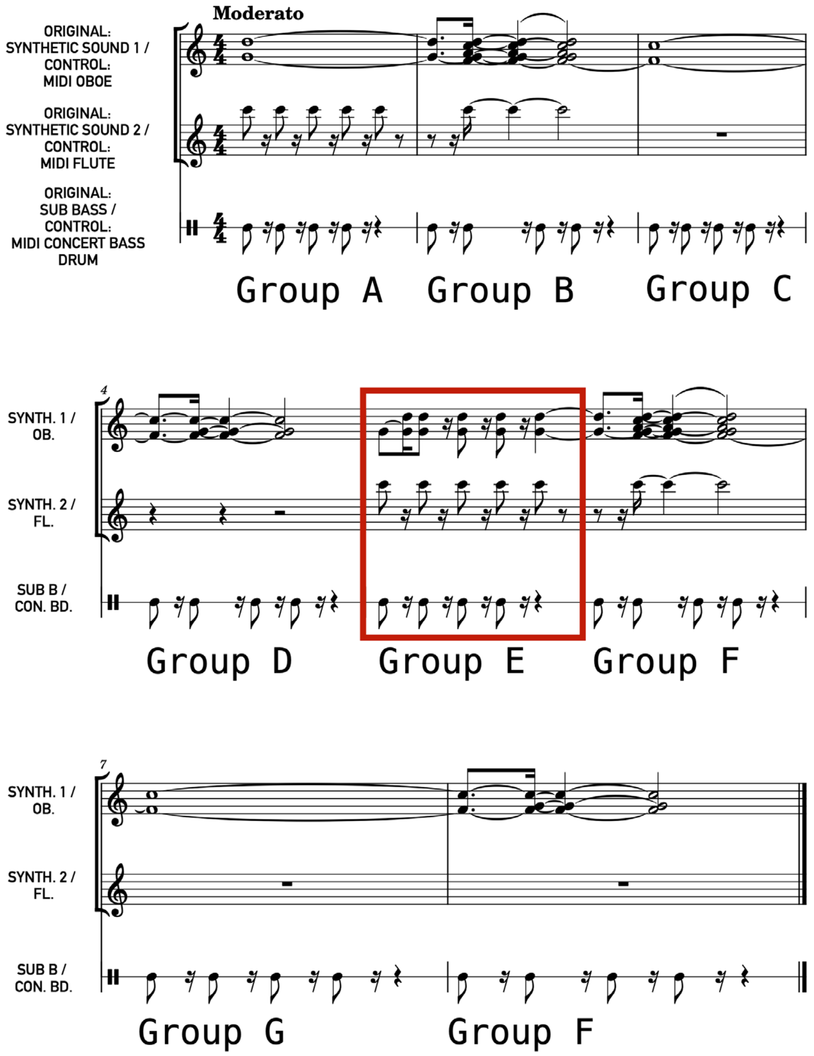


Figure 4: Approximate score transcription of the "Bass Beat Section" from Aphex Twin's "Bucephalus Bouncing Ball", with emphasis on b. 5, which exhibits rhythmic distortion of the main bass beat group. The instruments specified as "control" refer to a simple MIDI rendering of the track in *Garageband*, which is used for comparative analysis.

Animation 2: Parallel displayed animations of the "Bass Beat Section" from Aphex Twin's "Bucephalus Bouncing Ball" with *AudioVisualizer*'s "DotsAndLines" profile in both the original version and a Midi rendering based on the transcription shown as Fig. 4, in audiovisual matching and mismatching versions of both acoustic renderings.

Fig. 5 illustrates the cumulative vertical and horizontal pixel displacement data in correlation with Animation 2 and the two renditions of the analysed segment from "Bucephalus Bouncing Ball". The graph positioned atop Fig. 5 sequentially presents both renderings for comparison. Building upon earlier observations concerning the slope trends in cumulative vertical and horizontal pixel displacements, it is pertinent to note that the steeper slope evident in the data for the animation of the original track signifies a clearer delineation of beat groups and, more broadly, denotes a heightened degree of dynamism and motion plasticity compared to the MIDI rendering. While possibly discernible from both auditory data and the direct audiovisual experience of the renditions, Fig. 5 provides a distinctive quantitative perspective on the discussed segment through the lens of a particular animation mode.

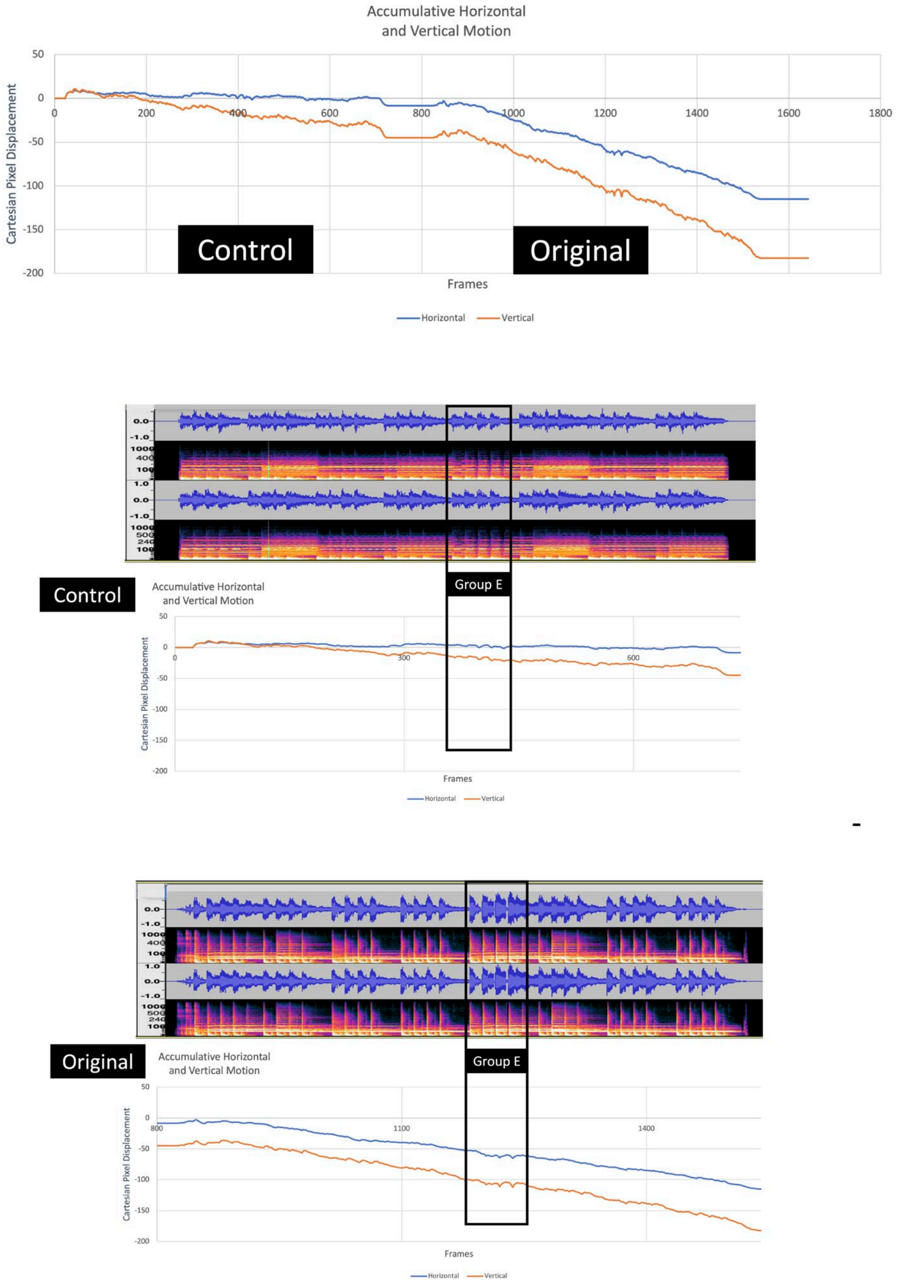


Figure 5: Cumulative optical flow data of both vertical and horizontal pixel displacement data with regard to Animation 2 and its renderings of both the original and MIDI versions of the discussed passage of “Bucephalus Bouncing Ball”.

It is noteworthy that the transpositional journey from the original sound through score transcription to MIDI rendering and abstract animation results in an animation image roughly isomorphic to the original case. While one might argue that the “mismatched” scenario of the original track and MIDI animation better supports the beat groups acoustically, such a preference for specific audiovisual alignment could be challenged in light of optical flow data indicating a more subdued occurrence of beat sequences in the MIDI rendering. Although the process of score transcription and MIDI control performance does not inherently prompt aesthetic appreciation, my comparative analysis suggests independent insights derived from optical flow analysis, which may uncover aesthetic nuances that could otherwise elude perception.

Notably, Fig. 5 accentuates again beat group E (refer to Fig. 4 for score information) as an intriguing case for comparison. What matters here is not only the difference between the original and control versions of the analysed passage in their shaping of the inherent beat groups, but also the distinctive parallelism between both vertical and horizontal pixel displacements in both versions that indicate a uniform motion of the animated "garland" in this particular animation modus. Indeed, as outlined in the score transcription of Fig. 4, the discussed passage of "Bucephalus Bouncing Ball" operates in the juxtaposition of three musical layers (Flute, Oboe, and Bass Drum in the applied MIDI control performance). Optical flow analysis of section E suggests an integrated uniform motion at this point, a feature that obviously could easily also be discovered through audiovisual listening. However, it is here shown to be discoverable through genuine quantitative analysis of a dynamic visual shape, indicating that the method of combining abstract animation with optical flow analysis has the potential for the detection of distinct candidate features of musical expression and motion that are inaccessible through acoustic data alone and, subsequently, may connect more directly to the aesthetics of music.

To underscore the complexities of beat group E in direct contrast to the remaining seven beat sequences and to expand the assessment of abstract music animation through optical flow data, in the remainder of this section I shall briefly revert to the original version of the track. Fig. 6 presents non-cumulative vertical motion in reverse order, synchronised with the bass beats of the analysed passage in terms of downward acceleration. However, optical flow analysis in this context encompasses three distinct regions of the garland-like animation shape (see Fig. 7 for definition), facilitating a more nuanced examination of the dynamic interplay among the track's layers, as illustrated in Fig. 4.

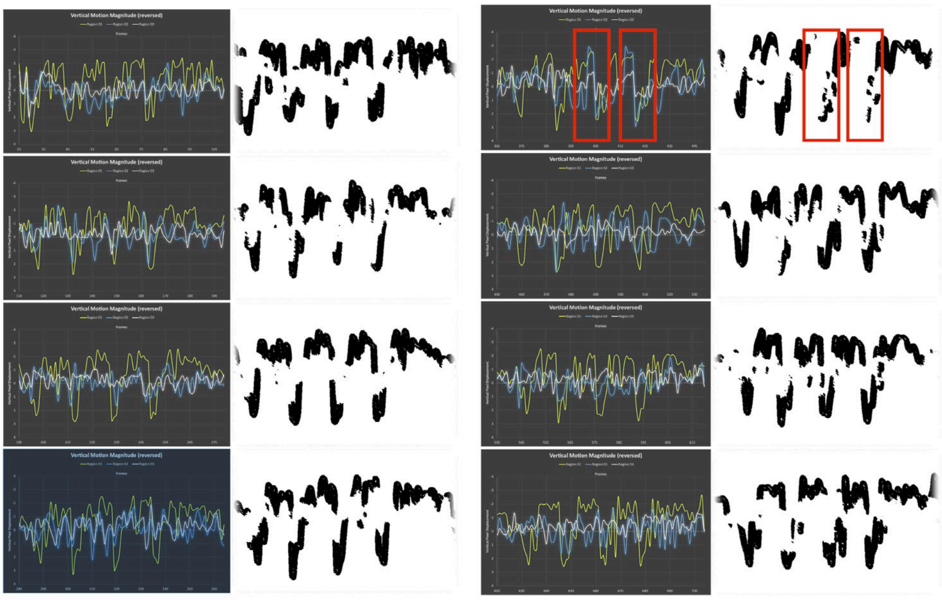


Figure 6a/6b: Vertical optical flow motion data extracted from three specific regions of the animation discussed hitherto (see Fig. 7 for details regarding region scope), portraying the “Bass Beat Section” from Aphex Twin’s “Bucephalus Bouncing Ball”. The *AudioVisualizer*’s “DotsAndLines” profile (see Fig. 1) is applied to each beat group within this passage (see Fig. 4 for score information). Additionally, an artistic filter overlay is employed to enhance visualisation, emphasising crucial differences in motion integration across the beat groups.

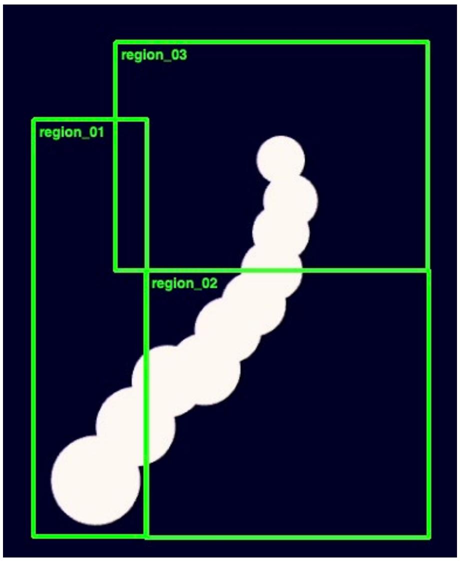


Figure 7: The three regions for optical flow analysis that underpin the analysis shown in Fig. 6.

Fig. 6 displays the versatility of optical flow analysis applied to spatialised musical spectra through animation. However, it also underscores the persistent challenges inherent in interpreting the copious amounts of data generated. In this instance, Fig. 6 employs an artistic filter to provide a coarse yet more readily understandable analogue image of the intricate data within specific regions. The application of this artistic filter results in a “whitewashing” effect, offering a rough indication of uniformity of motion across the three analysed regions. Group E displays the most distinct motion alignment among the three analysed regions due to its rhythmic characteristics, exhibiting a uniform spatial displacement that cannot be derived from acoustic analysis alone. However, noticeable differences among the other groups indicate that each beat group, despite superficial similarities, has its own unique shape and integration within the overall movement layers of the composition.

Different animation profiles lead to varied emergent properties and movements, all rooted in the objective acoustic spectrum. Hence, optical flow analysis provides a quantifiable means to explore how such mappings may influence the constitution of diverse forms of music perception, offering insights into the nuanced interplay between subjective interpretation and objective musical structure, which I shall discuss further in my second case study.

3 Frédéric Chopin's Prelude op. 28 no. 3 in Three Interpretations

Chopin's renowned Op. 28, a collection of 24 Preludes for piano, traverses all major and minor keys, presenting romantic miniatures characterised by a diverse array of textures, movements, and emotions. The third piece of the cycle, set in G Major, juxtaposes a buoyant ("leggiermente", yet intricate, wave-like accompaniment in the left hand with a sweeping, partially fanfare-like melody in the right hand, culminating in a unified passage where both hands converge in unison (see Fig. 9). This interplay between two dimensions opens up an intriguing space for interpretation in terms of dynamic, textural and kinetic balancing, integration and accentuation.

Applying a new animation setting in *AudioVisualizer* to its "Worm" scene (see Fig. 8 for details and Moshammer, 2023, p. 18, for context), this section discusses three different interpretations of Chopin's prelude in terms of optical flow, performed by Maria João Pires,[5] Ivo Pogorelich[6] and Seong-Jin Cho.[7]

Animation 3 illustrates the interpretations using the selected animation profile, while Animation 4 specifically focuses on the finishing sequence of the piece (bb. 28–32) at half speed. This approach highlights the fluctuations occurring in the centre of the animated galaxy. Fig. 10 assesses the three audiovisual representations with an emphasis on overall motion as analysed through optical flow. It compares the movement of the entire galaxy with the motion within its central region, as depicted in Fig. 11. The graphs in Fig. 10 are colour-coded to align with the structure of the piece's score, also illustrated in Fig. 9.

The employed animation profile, characterised by dynamic pushes (refer to Moshammer (2023) for further details), determines rotation speed based on dynamic fluctuations. In the context of Chopin's Prelude op. 28 no. 3, this method varies in its treatment of the two musical threads represented by the left and right hands. It captures overall dynamic changes in terms of rotation speed while also reflecting pitch and agogic changes through textural vibrations among the dots that form the galaxy.

Fig. 10 highlights bb. 28 and 29 in Pogorelich's interpretation (red arrow), revealing an intriguing asymmetry between the overall dynamics and textural fluctuations captured at the galaxy's centre. The details of these differences are difficult to detect through mere audiovisual

---

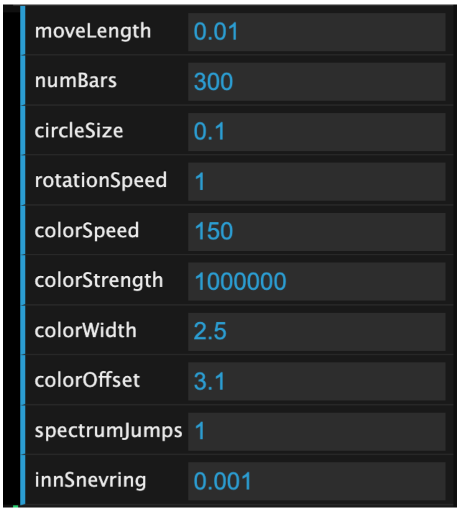
[5] From: Maria João Pires (piano), Royal Philharmonic Orchestra, & André Previn. (1994, July 5).

[6] From: Pogorelich, I. (1990, February 9). Frédéric Chopin: 24 Préludes op. 28. Deutsche Grammophon.

[7] Chopin Institute (2015). Seong-Jin Cho – Prelude in G major Op. 28 No. 3 (XVII International Chopin Piano Competition, third stage). Retrieved from https://www.youtube.com/watch?v=mStlxA97i_o.

listening and are not immediately apparent from the acoustic data alone (see Fig. 13 for a depiction of waveform and spectrum). This underscores the advantages of optical flow analysis in the study of abstract music animation.

Pogorelich's distinctive wave-like interpretation of bb. 28 and 29 appears to stem from his irregular timing of the passage. This interpretation is clearly audible in the half-speed version of Animation 4, where the changing frequency results in specific accelerations and decelerations of the repeated motif. This rendering is less influenced by overall dynamic changes compared to the interpretations of Cho and Pires.



moveLength	0.01
numBars	300
circleSize	0.1
rotationSpeed	1
colorSpeed	150
colorStrength	1000000
colorWidth	2.5
colorOffset	3.1
spectrumJumps	1
innSnevring	0.001

Figure 8: Settings in *AudioVisualizer*'s "Worm" profile as used in Animation 3.

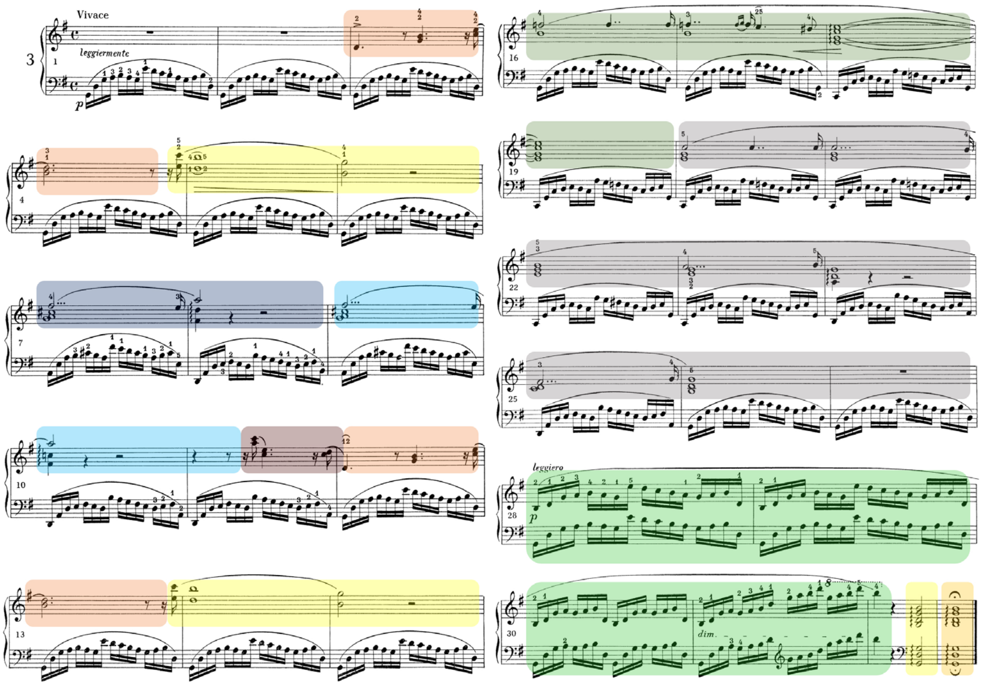


Figure 9a/9b: Annotated Score of Frédéric Chopin‘s Prelude op. 28/3[8]

Animation 3: Three interpretations of Frédéric Chopin’s Prelude op. 28 no. 3, performed by Maria João Pires, Ivo Pogorelich and Seong-Jin Cho, animated with *AudioVisualizer*’s “Worm” profile (see Fig. 8 for details).

Animation 4: An excerpt of Animation 3, highlighting bb. 28–32 in Chopin’s Prelude in the performances of Maria João Pires, Ivo Pogorelich and Seong-Jin Cho, played at half speed and with an emphasis on the moving galaxy’s centre.

---

[8] From: Chopin Complete Works Vol. I (Ignacy Jan Paderewski, ed.). Krakow: PWM, 1949. Retrieved from https://imslp.org/wiki/Preludes,_Op.28_(Chopin,_Fr%C3%A9d%C3%A9ric).

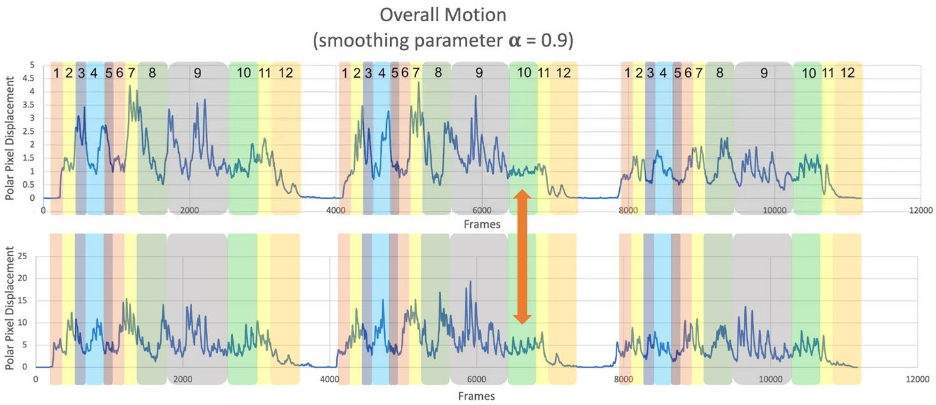


Figure 10: Overall motion magnitude data for Animation 3 in three graph regions, each representing the interpretations of Chopin's Prelude op. 28 no. 3 by Maria João Pires, Ivo Pogorelich, and Seong-Jin Cho, respectively (top graph). This is contrasted with the optical flow data for the galaxy's centre, shown in the region depicted in Fig. 11 (bottom graph). The colour coding corresponds to the annotated score of the piece as illustrated in Fig. 9.

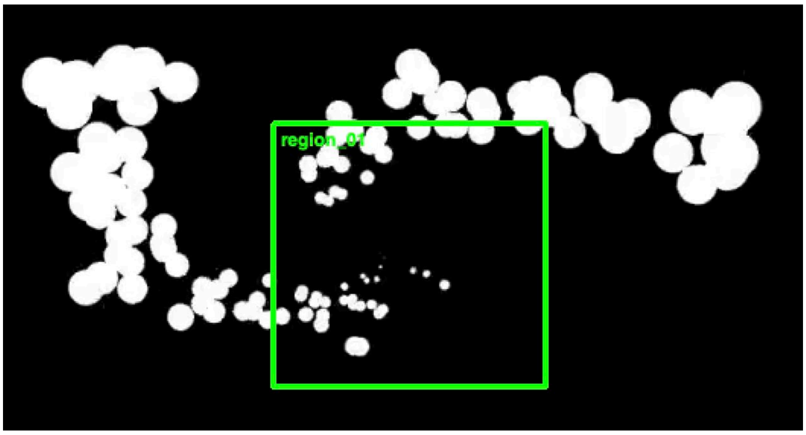


Figure 11: Optical flow region used in Fig. 10's bottom graph for data comparison.

In interpreting Chopin's Prelude op. 28 no. 3 from both pianistic and perceptual perspectives, several fundamental aesthetic factors come into play. Firstly, the degree of dynamic contrast between the left and right hand, along with the shaping of the melody in the right hand, is of utmost importance. Secondly, the precise articulation of the left-hand patterns, often played non-legato for clarity and transparency, significantly contributes to the piece's aesthetic impact. Lastly, the strategic (non-)use of the sustain pedal, capable of supporting legato articulation and dynamic shaping, influences the perceived "figure-ground" relationship between the right and left hands, carving the relief-like sound image that certain interpretations can evoke from the score.

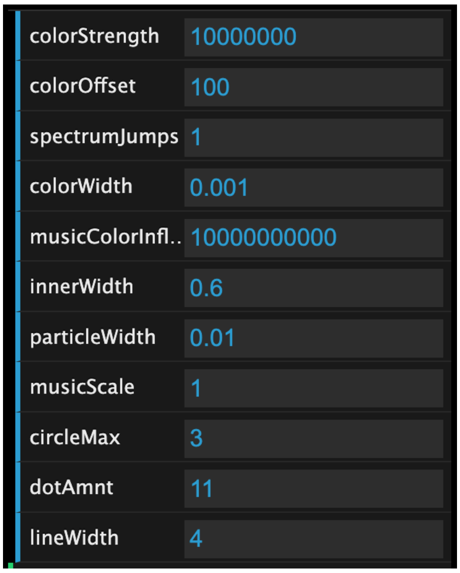
Even a cursory examination of Fig. 10 reveals that, in terms of motion distribution, the interpretations of Pires and Pogorelich align more closely with each other than either does with Cho's rendering. However, a more detailed comparison of the three graphs unveils significant differences among all three versions. To illustrate some of the insights that basic optical flow analysis can provide, let me address two specific excerpts in greater depth.

Firstly, concerning musical motion and the interplay between the two melodic and textural threads performed by each hand, the entrance of the first dotted motive by the right hand (first coloured graph sector in Fig. 10) exemplifies distinct motion profiles in the three interpretations (given the specific animation profile underlying this analysis). While a thorough examination may benefit from a return to the unsmoothed raw data, the graphs in Fig. 10 unmistakably demonstrate how Pires' rendition seamlessly integrates the initiation of the piece's first phrase with the motion of the introductory solo left-hand patterns (preceding the first coloured segment in Fig. 10 and setting the "ground speed" of the animation's rotation). In Pires' performance, acceleration commences after the ascending dotted motive enters. Conversely, Pogorelich's rendition, depicted in the analysis presented by the middle graph of Fig. 10, accelerates the musical progression immediately upon the right hand's entrance, relegating the left hand to the background right from the outset. Cho's interpretation, as indicated in the final graph of Fig. 10, also imbues the right hand's initiation with elevation and dynamism, albeit in a more subdued manner compared to Pogorelich, before accelerating towards the peak of the phrase.

A second noteworthy instance deserving closer scrutiny is the recurrence of the initial motive in bb. 12–15 (coloured graph sectors 6 and 7 in Fig. 10) of Chopin's prelude, which is foreshadowed by a new dotted motive in b. 11 (coloured sector 5). Both Pires and Pogorelich infuse the preparation for the re-entry of the initial phrase with heightened dynamism, signifying, in terms of the applied animation profile, the highest degree of acceleration in these renderings (which Pires maintains more consistently, while Pogorelich accentuates the dotted rhythm to a greater dynamic extent). Conversely, Cho, while similarly accelerating the motive's repetition in bb. 12–15 with heightened dynamic, imbues the subsequent section with an actual increase in dynamism, marking the highest degree of motion in his performance if perceived in terms of Animation 3.

Describing and annotating the graphs in Fig. 10 in detail presents a significant challenge due to the subtle differences observed among the three interpretations under discussion. However, this challenge does not indicate a theoretical shortcoming in itself, but rather provides an objective basis for understanding the nuanced level at which music performance and appreciation may operate. While both the overall motion distribution of the moving galaxy and its inner textural movement are direct results of mapping acoustic frequency and amplitude, the interplay between the animated image and various levels of optical flow analysis opens a novel window into a discussion of aesthetically relevant differences among the interpretations of Chopin's piece.

This complexity is further illustrated by the subsequent animation profile, which is based once again on *AudioVisualizer*'s "DotsAndLines" scene. Implemented in Animation 5, this new animation mode enables a more seamless integration of the musical streams of both the left- and right-hand parts in Chopin's prelude. This stands in contrast to the previous approach, where the vibrancy of the left hand was somewhat obscured within the moving "galaxy" depicted in Animation 3.



colorStrength	10000000
colorOffset	100
spectrumJumps	1
colorWidth	0.001
musicColorInfl..	10000000000
innerWidth	0.6
particleWidth	0.01
musicScale	1
circleMax	3
dotAmnt	11
lineWidth	4

Figure 12: Settings in *AudioVisualizer*'s "DotsAndLines" profile as used in Animation 5.

Animation 5: Three interpretations of Frédéric Chopin's prelude op. 28 no. 3, performed by Maria João Pires, Ivo Pogorelich and Seong-Jin Cho, animated with *AudioVisualizer*'s "DotsAndLines" profile (see Fig. 12 for details).

Fig. 13 offers a detailed analysis and comparison of three regions within the animated shape, introducing complexity to the examination. The wealth of optical flow data generated through this analysis underscores once more the intricate nature of this method, raising questions about the extent to which the human mind comprehends musical texture, motion, and expression in relation to dynamic and temporal elements. As shown in Fig. 13, smoothing the optical flow data can act as a cipher for the gestalt-based perceptual synthesis of music's acoustic flow. For example, the graph with 0.995 alpha data reduction in Fig. 13 highlights two noticeable features of the optical flow analysis of the three animation regions.

Alpha data reduction, as referenced in Fig. 13, refers to the application of exponential smoothing to the optical flow data using a smoothing constant, or alpha, set at 0.995. In exponential smoothing, the alpha (α) value determines the weight given to the most recent data points versus past values. An alpha close to 1.0 (such as 0.995) gives very high weight to recent observations, resulting in a smoothed curve that still retains nearly all the dynamic fluctuations of the raw data, but with minor noise suppressed. In the context of Fig. 13, this smoothing technique allows for a clearer visual interpretation of motion trends across the three regions of the animated shape, without discarding temporal complexity. The high alpha thus preserves most of the original signal while reducing jitter, making the flow patterns more perceptually coherent. This method of data reduction supports the analogy of optical flow functioning as a cipher for how the mind integrates audio-visual motion into an expressive musical percept.

The first graph segment, representing Pires' performance, indicates heightened motion in region three of the optical flow analysis applied here during the recurrence of the piece's initial phrase in b. 13 that has already been discussed in relation to Animation 3. However, in this case involving the analysis of three animation regions, one can deduce additional musical information from optical flow analysis. It is crucial to observe that the three analysed regions, representing the lower, middle, and higher sections of the animated mirrored C-shape, correlate with the lower, middle, and higher bands of the music spectrum. Hence, the indicated passage in Fig. 13 where the third region becomes relatively accelerated not only indicates heightened dynamics but also an increased usage of the sustain pedal that creates a thicker acoustic wave in the left hand at this point, dynamically expanding the extension of the mirrored C-shape by equally reducing the fluctuation in the animation's top and bottom regions. Such analysis introduces levels of scrutiny that surpass the reading of acoustic data due to the import of dynamic spatialisation. Specifically, while in Fig 13, the analysis is broken down into three regions of the animated shape, the interpretive task here entails an integrative reading that reflects on the overall shaping of expression. It is also worth noting that critical aspects of music perception, such as modelling musical contour and melody based on pitch qualities alone, cannot be captured by the method presented here, despite its importance in analysing the overall imprint music perception can generate.

Fig. 13 also indicates the relatively low motion magnitude in the third analysed animation region concerning Cho's rather subdued performance of Chopin's prelude (refer to the Fig. 13's third and final graph sector in which the arrow indicates the happening in b. 30). Notably, perceived motion and textural vibrancy do not necessarily correlate with motion magnitude but rather express themselves in the degree of change, which can, for instance, be facilitated by mere pitch alterations. However, the highlighted passage in Cho's interpretation provides evidence for a distinctive motion formation within the animated mirrored C-shape that the two versions it is compared with do not exhibit. Basically, Cho reduces activity in the lower spectrum of the left hand here, subsequently, pushing activity of the mirrored C-shape upwards. Describing such nuances is a delicate task, akin to poetry rather than science, leading one to the discourse in Philosophy concerning "non-conceptual content" (see Luntley, 2003; Peacocke, 2001) or "undefinable" qualia (see Schiavio & van der Schyff, 2016). However, while optical flow analysis does not "semanticise" music's dynamism and expressivity, which would require an additional annotation system, it offers, as demonstrated here, measurements that elucidate possibly important expressive distinctions.

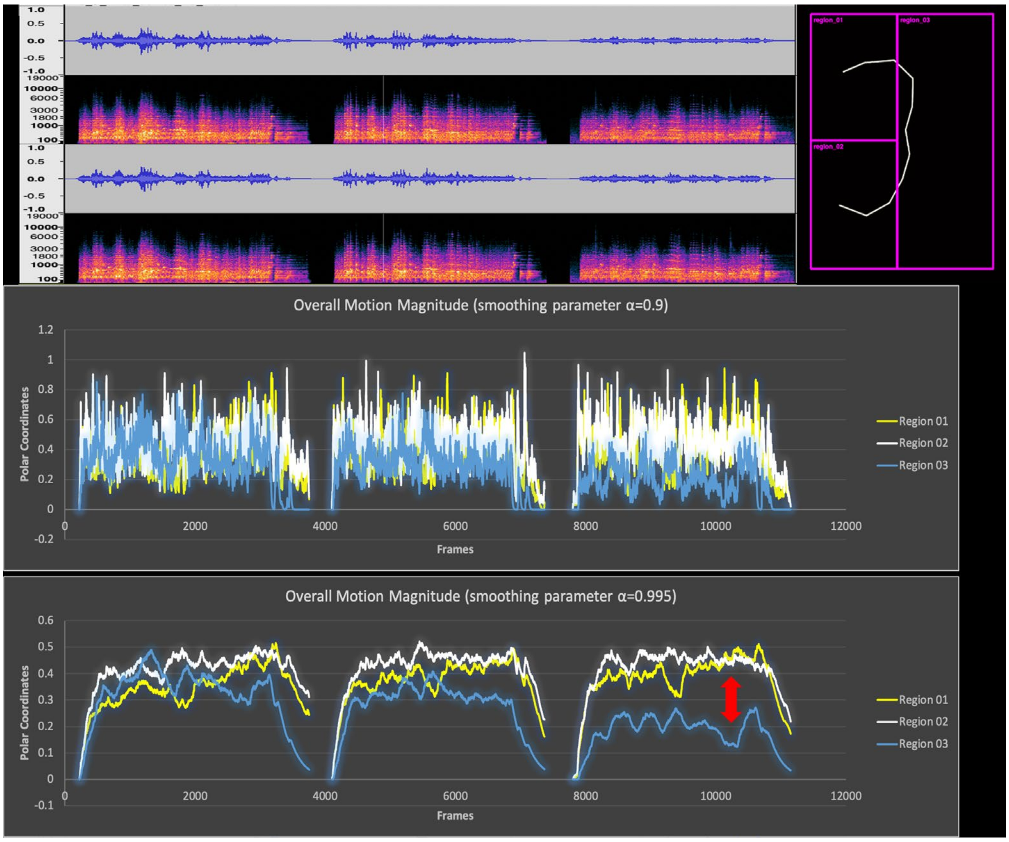


Figure 13: Two series of three graph regions each, illustrating interpretations of Chopin's Prelude op. 28 no. 3 by Maria João Pires, Ivo Pogorelich, and Seong-Jin Cho, respectively. Each series represents varying degrees of smoothing and is juxtaposed with the piece's acoustic data. Furthermore, three regions of optical flow analysis are incorporated for comprehensive comparison.

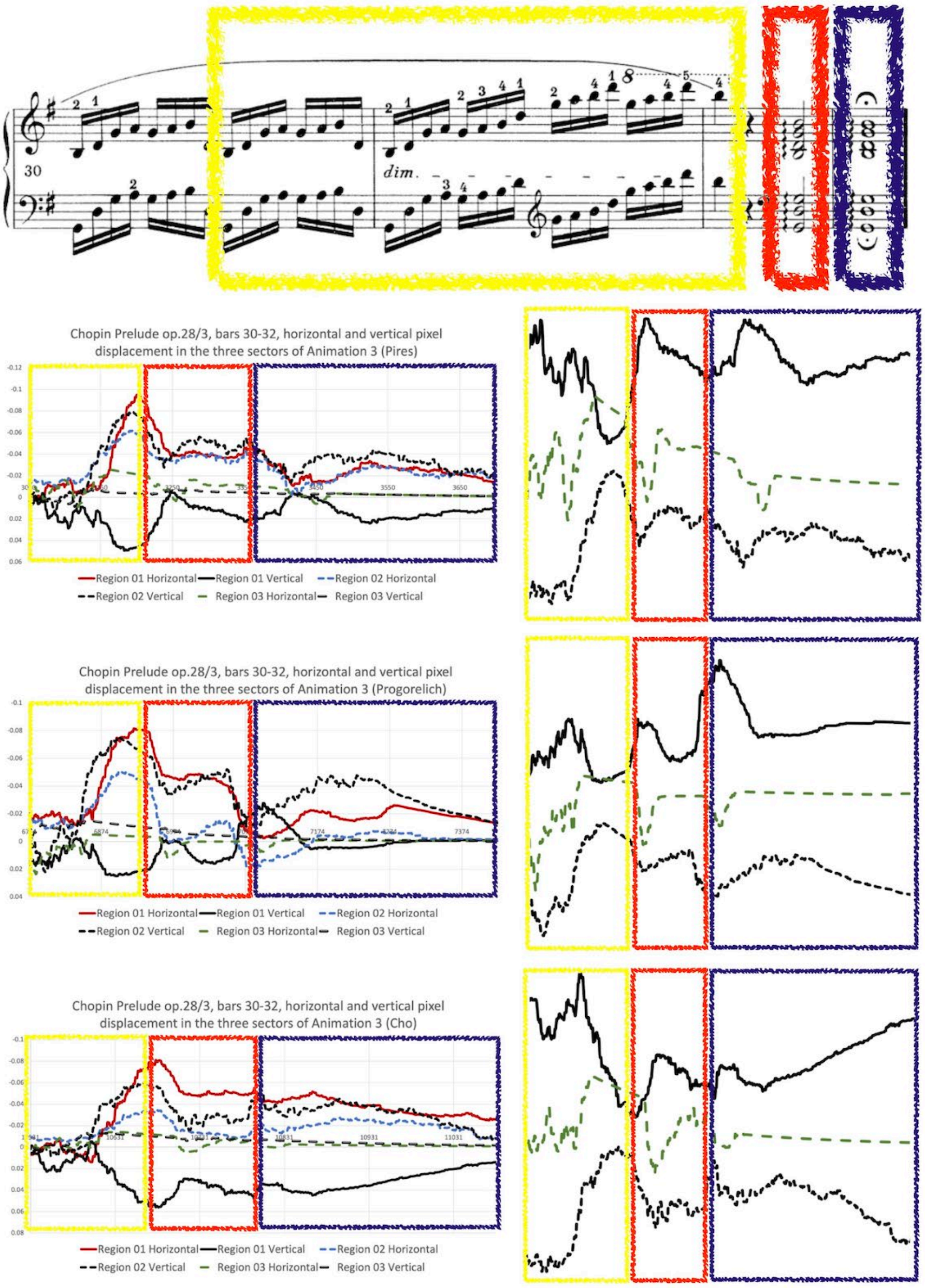


Figure 14: Vertical and horizontal pixel displacement data for the ending of Chopin's Prelude op. 28 no. 3 (bb. 31 and 32) of Animation 5. The graph pairs represent the performances of Maria João Pires, Ivo Pogorelich and Seong-Jin Cho, respectively (see text for more details regarding the reading of the graphs).

Finally, concerning the conclusion of Chopin's Prelude op. 28 no. 3, I would like to offer a brief reflection on potential pathways for a detailed analysis of micro-expressions and motion in the context of optical flow analysis applied to abstract music animation. In Fig. 14, the initial graph of each pair depicts all flow data regarding vertical and horizontal movement for the three animation regions, as illustrated in Fig. 13, with the order reversed to facilitate a more intuitive interpretation of directionality. The accompanying graphs to the right of each pair rearrange a subset of three data series, presenting an analogue representation that tracks the animation's "opening" and "closing" of the animated C-shape, with the upper and lower lines in the graph representation. The central data series pertains to horizontal movement to the right, where “higher” values indicate a greater extension to the right. From these graphs, one can readily infer the movement of the animated C-shape. It is noteworthy that while the trajectories of motion may vary slightly, all three interpretations feature a “negative” climax characterised by a contracting motion following the upward semiquaver run (b. 31). However, significant differences, once again challenging to articulate unequivocally, emerge among the three performances in the evolving expansion of the piece’s final two chords, which the graph information in Fig. 14 demonstrates in great detail.

In facilitating such detailed reconstruction and quantitative evaluation of gestural emergent shaping in idiosyncratic animated mappings of musical spectra, optical flow analysis enables the precise extraction of a variety of micro-expressive elements. This method goes beyond traditional acoustic analysis by translating the animated shapes, such as a C-shape that opens according to acoustic data, into numerical graph information. This numerical representation directly corresponds to the analogue geometrised image derived from the musical spectrum in visualisation, capturing subtle gestural nuances that contribute to the fine-tuned appreciation of music. Through this advanced analytical approach, one can more comprehensively understand the intricate motion patterns and expressive dynamics that play a critical role in the perception and appreciation of musical expression.

4 Discussion

Optical flow analysis emerges as a potent tool for quantitatively zooming-in on the dynamics of music animation. In instances where music animation is algorithmically generated, such as with *AudioVisualizer*, it adheres to a distinct geometrical rendering of the musical spectrum. By methodically examining the spatial arrangement of musical data, optical flow analysis introduces a novel lens for apprehending the dynamic and expressive facets of music.

One of the pivotal advantages of optical flow analysis in automated abstract music animation lies in its capacity to transcend subjective audiovisual experiences. Unlike conventional approaches, optical flow analysis has the capability to discern unique features and interpretations of music that may elude detection through auditory or visual means alone. Furthermore, the method provides insights into the intricacies of music listening, revealing possible discrepancies in interpretation and potentials for mental shaping, by systematically scrutinising the musical spectrum from a transformative standpoint that yields dynamic visual representations.

While audiovisual music rendering may enhance educational contexts by priming perception, it is paramount to recognise that optical flow analysis primarily functions as an extension of acoustic data rather than seeking to serve as an independent aesthetic contribution in terms of supporting audiovisual experience. Through diverse techniques such as capturing optical flow in polar and Cartesian coordinates and focusing on specific regions of animated shapes, optical flow analysis facilitates a comprehensive exploration of the musical spectrum in terms of motion, expression, and emergent musical gestalts. This meticulous screening potentially transcends both the fidelity of listening and a straightforward account of acoustic data, as evidenced in the analysis sections of this article.

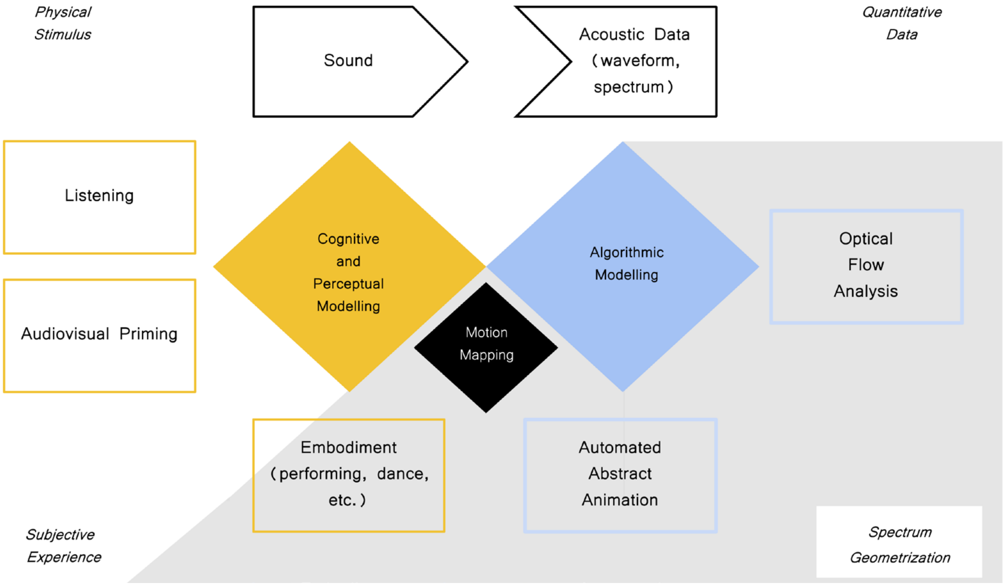


Figure 15: A theoretical framework of musical motion and its modelling.

Fig. 15 juxtaposes human perception and cognitive models of music with algorithmic processing and the measurement of spectral data. Highlighting abstract motion and its optical flow in abstract animation, such a bird's-eye view also reveals a fundamental limitation: essential musical attributes like melody, harmony, and timbre are not directly captured. Nevertheless, by focusing on music's dynamic nature through motion, this approach paves the way for what could be termed "algorithmic hermeneutics". Coupled with a more traditional, holistic hermeneutics of music analysis and in conjunction with "audiovisual listening", this algorithmic approach may forge new pathways for comprehending specific pieces of music as well as a deeper theoretical understanding of what musical expression actually is.

Integrating algorithm-based abstract music animation in terms of a geometrisation of the musical spectrum, optical flow analysis directs theoretical attention towards the microscopic dynamic and kinetic layers of music's expressive functionality, rather than attempting to categorise music under broad emotional labels. The quantitative nature of this method strives to extract the myriad subtleties of acoustic sensemaking from the realm of the subjective and ineffable, bringing to light micro-gestures and micro-motions that may be overlooked under the guise of subjectivity or disregarded scientifically altogether.

Recent research delves into the fields of data sonification (Hermann, Hunt, & Neuhoff, 2011), "algorithmic synaesthesia" (Sagiv, Dean, & Bailes, 2012), and sensory substitution (Lloyd-Esenkaya et al., 2020), emphasising the transferability of structural sensory information and the varied sensory modes of "sense-making" in response to isomorphic stimuli. Within the context of audiovisual music experiences, however, it is crucial to recognise the significant asymmetry between visual imagery and the multidimensional auditory experience. Music experience encompasses the simultaneous perception of polyphonic melody, harmony, rhythm, timbral qualities, textural associations, and expressive timing, making it unique and unmatched by visual perception in terms of integrative information processing and temporal resolution.

Subsequently, when considering abstract visual imaging, which does not directly evoke ecological scenery or real-life emotional expressions as cartoons do, one might question the relevance of engaging in audiovisual analysis within the realm of music hermeneutics. Despite these concerns, the geometrisation of musical spectra, addressing music's dynamic changes and emergent expressive tropes, reveals subtle aspects of musical texture and motion that might otherwise remain hidden. Optical flow analysis, in particular, enhances the examination of intricate details in these visual representations. This technique facilitates a sophisticated understanding of the interaction between different frequency components over time, providing a richer, multidimensional perspective on music analysis. The interplay between sound and visual elements can uncover deeper layers of meaning and structure within the musical composition, revealing insights that might not solely depend on auditory processing.

5 Conclusion

The considerations presented in this article give rise to three interconnected tasks and questions. One revolves around technology, the second around methodology, while the final cornerstone rests on a more fundamentally philosophical inquiry.

Within the realm of technology, particularly in light of recent strides in artificial intelligence and its pursuit of Artificial General Intelligence (AGI), there arises the possibility of envisioning a software package that not only amalgamates diverse modes of abstract animation, delving into the geometrisation of musical spectrum and optical flow analysis but also offers insights into musical features and characteristic performance parameters through an automated annotation system. However, articulating such a desideratum requires a comprehensive examination of the scope and limitations of what fundamentally must, at least preliminarily, be rooted in a trans-modal, albeit possibly algorithmically supported hermeneutics, as exemplified within this article.

From a purely algorithmic standpoint, one might question the necessity of invoking audiovisual imagery or graphical data presentations if both can ultimately be traced back to an acoustic analysis of musical spectra. However, eschewing the utilisation of music's dynamic audiovisual image would forfeit potential benefits in training, for instance, neural networks for feature detection based on visual input, besides taking away a crucial reference point for the aforementioned trans-modal hermeneutic task.

The application of optical flow analysis to abstract score animation has a microscopic effect, similar to running an animation in slow motion or simply reducing the speed of a recording. It opens a window to details and differences that not simply demand acknowledgment *sui generis* but for an assessment in terms of their structural, expressive, and emotional relevance in the overall appreciation of music. For instance, a versatile AGI application in the spirit of the considerations in this article may be able to lay a more objective and precise foundation for discussion in relation to the mental modelling concerning music.

Subsequently, the application of optical flow analysis to abstract animations derived from musical spectrum visualisation introduces layers of complexity that resonate with the mission the humanities in seeking to illuminate the unique significance and contextual relevance of individual artefacts. In this context, optical flow analysis offers a means to explore the intricacies of artistic expression, unveiling nuances in motion and emergent dynamics that contribute to the overall aesthetic experience. Embracing such complexities not only enriches our comprehension of music as an art form but also underscores the multifaceted nature of human creativity and perception. Ultimately, such a program could pose concrete challenges to the classical body/mind problem when viewed through a reductionist lens, ironically facilitated by the application of tools provided by artificial intelligence.

References

Ammirante, P., Patel, A. D., & Russo, F. A. (2016). Synchronizing to auditory and tactile metronomes: A test of the auditory-motor enhancement hypothesis. *Psychonomic Bulletin & Review*, *23*, 1882–1890.

Cancino-Chacón, C. E., Grachten, M., Goebl, W., & Widmer, G. (2018). Computational models of expressive music performance: A comprehensive and critical review. *Frontiers in Digital Humanities*, *5*(25). https://doi.org/10.3389/fdigh.2018.00025.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477.

Cook, Nicholas. (2013). *Beyond the score: Music as performance*. Oxford University Press.

Cox, C., & Warner, D. (Eds.). (2004). *Audio culture: Readings in modern music*. Continuum International Publishing Group.

DeNora, T. (2016). *Music asylums: Wellbeing through music in everyday life*. Routledge.

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 1–23.

Fischer, T., & Brox, T. (2015). Optical flow estimation. In K. Ikeuchi (Ed.) *Computer vision: A reference guide* (pp. 391–410). Springer.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836.

Gabrielsson, A., & Lindström, E. (2010). The role of structure in the musical expression of emotions. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 367–400). Oxford University Press.

Gaver, W. W., & Norman, D. A. (1988). *Everyday listening and auditory icons* [unpublished doctoral dissertation]. University of California.

Gescheider, G. A. (2013). *Psychophysics: The fundamentals* (3rd ed.). Psychology Press.

Gofton, L. (2017). *Meme, counter-meme: From cyberculture to generations*. Springer.

Goodman, N. (1976). Languages of art: An approach to a theory of symbols. Hackett publishing.

Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics*, *72*(3), 561–582.

Guizilini, V., Lee, K.-H., Ambruş, R., & Gaidon, A. (2022). Learning optical flow, depth, and scene flow without real-world labels. *IEEE Robotics and Automation Letters*, *7*(2), 3491-3498. https://doi.org/10.1109/LRA.2022.3145057.

Habibi, A., Cahn, B. R., Damasio, A., & Damasio, H. (2016). Neural correlates of accelerated auditory processing in children engaged in music training. *Developmental Cognitive Neuroscience*, *21*, 1–14.

Hall, K. C., Smith, H., McMullin, K., Allen, B., & Yamane, N. (2017). Using optical flow analysis on ultrasound of the tongue to examine phonological relationships. *Canadian Acoustics*, *45(1)*, 15–24.

Hermann, T., Hunt, A., & Neuhoff, J. G. (2011). *The sonification handbook* (Vol. 1). Berlin, Germany: Logos Verlag.

Hohwy, J. (2013). *The Predictive mind*. Oxford University Press.

Iordanescu, L., Grabowecky, M., & Suzuki, S. (2013). Action enhances auditory but not visual temporal sensitivity. *Psychonomic Bulletin & Review*, *20*, 108–114.

Jan, S. (2007). *The memetics of music: A neo-Darwinian view of musical structure and culture*. Routledge.

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Physics of Life Reviews*, *10*(3), 235–266.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*(5), 559–621.

Kahn, D. (1999). *Noise, water, meat: A history of sound in the arts*. MIT Press.

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience*, *15*(3), 170–180.

Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology* (Rev. ed.). New York, NY: Liveright.

Kraus, N., Slater, J., Thompson, E. C., Hornickel, J., Strait, D. L., Nicol, T., & White-Schwoch, T. (2014). Music enrichment programs improve the neural encoding of speech in at-risk children. *Journal of Neuroscience*, *34*(36), 11913–11918.

Levitin, D. J. (2006). *This is your brain on music: The science of a human obsession*. Penguin.

Liu, Y., Alizadeh, D., & Tang, T. K. (2019). Optical-flow analysis for noninvasive assessment of tumor microenvironment. *IEEE Transactions on Biomedical Engineering*, *66*(8), 2293–2300.

Lloyd-Esenkaya, T., Lloyd-Esenkaya, V., O'Neill, E., & Proulx, M. J. (2020). Multisensory inclusive design with sensory substitution. *Cognitive Research: Principles and Implications,* *5*(1), 37. https://doi.org/10.1186/s41235-020-00240-7.

Luntley, M. (2003). Non-conceptual content and the sound of music. *Mind & Language*, *18*(4), 402–426.

Luo, W., & Ning, B. (2022). High-dynamic dance motion recognition method based on video visual analysis. *Scientific Programming*, 2022(6724892,), 1–9. https://doi.org/10.1155/2022/6724892.

Margulis, E. H. (2013). *On repeat: How music plays the mind*. Oxford University Press.

Moshammer, G. (2023). Abstract animation, emergent audiovisual motion and micro-expression: A case study of analogue music tracking with Robert Schumann's forest scenes in AudioVisualizer. *Journal of Creative Music Systems*, *7*(1), 1–23. https://www.jcms.org.uk/article/id/1311/.

Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R., & Ullen, F. (2014). Practice does not make perfect: No causal effect of music practice on music ability. *Psychological Science*, *25*(9), 1795–1803.

Nettl, B. (2005). *The study of ethnomusicology: Thirty-one issues and concepts*. University of Illinois Press.

Niehorster, D. C. (2021). Optic flow: A history. *I-Perception*, *12*(6), 1–49. https://doi.org/10.1177/20416695211055766.

Peacocke, C. (2001). Does perception have a nonconceptual content? *The Journal of Philosophy*, *98*(5), 239–264.

Peretz, I., Vuvan, D. T., Lagrois, M. É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140090.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, *41*(1), 245–255.

Repp, B. H. (1992). Music as motion: A synopsis of Alexander Truslit's (1938) *Gestaltung und Bewegung in der Musik*. Haskins Laboratories Status Report on Speech Research. SR-111/112, 265–278.

Repp, B. H., & Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological Research*, *68*, 252–270.

Sagiv, N., Dean, R. T., & Bailes, F. (2011). Algorithmic synesthesia. In R. T. Dean (Ed.), *The Oxford handbook of computer music* (pp. 294–311). Oxford University Press.

Schaeffer, P. (2012). *In search of concrete music* (C. North and J. Dack, Trans.). University of California Press. (Original work published 1952).

Scherer, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research*, *33*(3), 239–251.

Schiavio, A., & van der Schyff, D. (2016). Beyond musical qualia. Reflecting on the concept of experience. *Psychomusicology: Music, Mind, and Brain*, *26*(4), 366–378.

Strait, D. L., & Kraus, N. (2011). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00113.

Tagg, P. (2012). *Music's meanings: A modern musicology for non-musos* (Vol. 5). Mass Media Music Scholars' Press.

Temperley, D. (2019). *The cognition of basic musical structures*. MIT Press.

Titon, J. T., Cooley, T. J., Locke, D., McAllester, D. P., & Rasmussen, A. K. (2008). *Worlds of music: An introduction to the music of the world's peoples* (5th illustrated ed.). Belmont, CA: Cengage Learning.

Visi, F., Coorevits, E., Schramm, R., & Miranda, E. R. (2017). Musical instruments, body movement, space, and motion data: Music as an emergent multimodal choreography. *Human Technology, 13*(1), 58–76.

Wertheimer, M. (1923). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–88). London, England: Routledge & Kegan Paul. (Reprinted from *Untersuchungen zur Lehre von der Gestalt II, Psychologische Forschung,* 4, 301–350.)

Wittgenstein, L. (1966). *Lectures & conversations on aesthetics, psychology, and religious belief.* University of California Press.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*(4), 420–422.

Xenakis, I. (1992). *Formalized music: Thought and mathematics in composition* (S. Kanach, Trans.; 2nd rev. English ed.). Pendragon Press.

Zendel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychology and Aging, 27*(2), 410–417. https://doi.org/10.1037/a0024816.