Abstract animation, emergent audiovisual motion and micro-expression: A case study of analogue music tracking with Robert Schumann's *Forest Scenes* in *AudioVisualizer*

Gerald Moshammer
Mahidol University International College, Thailand
gerald.mos@mahidol.edu

Abstract

Abstract animation in the form of "visual music" facilitates both discovery and priming of musical motion that synthesises diverse acoustic parameters. In this article, two scenes of *AudioVisualizer*, an open-source Chrome extension, are applied to the nine musical poems of Robert Schumann's *Forest Scenes*, with the goal to establish a basic framework of expressive cross-modal qualities that in audiovisual synchrony become apparent through visual abstraction and the emergence of defined dynamic *Gestalts*. The animations that build this article's core exemplify hands-on how particular ways of real-time analogue music tracking convert score structure and acoustic information into continuous dynamic images. The interplay between basic principles of information capture and concrete simulation in the processing of music provides one crucial entry point to fundamental questions as to how music generates meaning and non-acoustic signification. Additionally, the considerations in this article may motivate the creation of new stimuli in empirical music research as well as stimulate new approaches to the teaching of music.

Keywords: Abstract animation, cross-modal expression, audiovisual synchrony, music expression, musical motion, Robert Schumann.

1 Introduction

The world-renowned pianist Krystian Zimerman, in an interview with the BBC,[1] has something rather surprising to say regarding music perception:

"...I'm realising more and more that music is not an audio experience and the digital technique actually showed me this…it so clearly transmits the sounds that you can't hear the music anymore…and music is not sound. We are using the sound to create music but music is actually more organising people's emotions in time and it's more the time flow, the story you are telling…going by more and more perfect sound you're not necessarily achieving a better story…because there will be a lot of factors which will start to disturb the listener, the perfection of sounds that is kind of overexposing itself…in the last ten years I'm listening to the flow of music…if I really want to hear music I put it into my car and I drive around the house because then the conscious part of the brain is occupied with the road and the music goes right there where should go and I was always curious why is it that way, why do I hear so clearly what's wrong in the records when I'm in the car and you know the basic noise of the car is covering all the details; so I stop listening to details, my mind stops being distracted by the details, and I listen to that what in the music is the most important, telling a story…"

Zimerman's view on "musical storytelling" lends itself to juxtaposition with "structural listening" for which especially Theodor W. Adorno (1963) provided a strong case on its own. However, while attention to music's expressivity does by no means necessitate unawareness of mu-

[1] The interview aired on Saturday 10 May 2008, 12:15-13:00, on Radio 3. See https://www.bbc.co.uk/radio3/musicmatters/pip/lwjxu/. To the best of my knowledge, the interview is not available any more on the official BBC website, but still can be found on social media platforms such as YouTube.

sical structure, Zimerman's idiosyncratic ideas regarding music perception pose a more fundamental challenge to the "positivist" credo that characterises certain segments of more recent performance studies. As Goebl *et al.* (2014, p. 225) summarise Gabrielsson (2003): "The main issue is that of obtaining reliable measurements, for each performed tone, of parameters such as timing, amplitude, and pitch, which are the main attributes investigated in performance research". Scientific approaches to music, its performance and reception, seek reliable data, with acoustic analysis constituting the natural departure point. However, both the perception of music and the mental formation of performance goals are not explainable through acoustic information alone. Borrowing from John Searle (1980, p. 27), the issue basically boils down to the gap between physics and semantics. "Music is not sound", as Zimerman stipulates. He is even going a step further, speaking of the distraction (acoustic) details may cause, a thought that is curiously linked to another aspect in Zimerman's reflections, i.e., the telling of a story in terms of "organising emotions in time". What Zimerman is hinting at in his thought-provoking remarks appears to be the interplay between information reduction and the emergence of defined perceptual qualities. Such a "less is more", or, at least, "less is enough" principle is most directly linked to studies of the visual perception of biological motion in terms of light point displays. Basically, only a pointed sketch of visual contour is necessary in order to identify forms of human motion such as walking, running or stair climbing, a phenomenon that has also been studied in relation to emotion perception in dance (Dittrich *et al.*, 1996). However, anyone who has listened to music behind a closed door, or experienced music with ear plugs, can probably relate to the profound effects that reduced acoustic fidelity can also have on the perception of music, especially when it is performed with acoustic instruments.

As a temporal art form, music connects the acoustic past with the acoustic future, invoking both memory and prediction in the mental representation of meaningful units such as motifs, phrases and melodies. Musical movement is essential for the constitution of music's temporal *Gestalts*, making it an idiosyncratic case of motion interpretation. Generally, understanding the world around us requires a grasp of how things move, with the concept of motion falling under the idea of change. Change itself can be manifold and plays out in either logical, physical or phenomenal space. Music exemplifies this complexity with its intrinsic relationship to all three of the aforementioned dimensions. First, music can exhibit a syntactically well-defined structure that enters score notation, navigating through a logical space that feeds melodic, harmonic and rhythmic analysis. Second, music emerges from the sound spectrum of waves and literally energises physical space, causing the ear's tympanic membrane to vibrate, which initiates neural processing of the acoustic signal. Third, music only "exists" in hearing, captivating phenomenal space that the listener can feel and that unfolds on the backdrop of integrating mental activity through modes of association, imagination and subjective timing. The logical, physical and phenomenal layers of music's functioning are methodologically difficult to control and synthesise. Scientific approaches to music tend particularly to show interest in what biologically evokes and has evolutionarily sustained music, drawing upon measurable effects music is able to generate, be it through the regulation of emotions, therapeutic benefits or mere arousal for the sake of pleasure. As a genuine art form, however, music implicates more perceptual dimensions than a simple reward structure.

Because the scientific study of music perception and aesthetics is mainly committed to analytic bottom-up modelling, musical motion tends to be linked to the measurement and evaluation of music's continuous progression. In particular, the tendency to derive movement in music solely from tempo, musical metre and beat distribution and to interpret musical motion in terms of pulse (Repp, 1989), expressive timing (Repp, 1995) or basic physical motion such as in *ritardandi* (Todd, 1995) is proof of a widespread adherence to an analytical musicological model that has the tendency of overlooking motion perception's ability to synthesise musical layers. Honing (2005) suggests an alternative "perception-based" view that also acknowledges note density, rhythmic structure and global tempo instead of adhering to basic kinematic principles in the explanation and prediction of expressive musical timing. Alternative approaches to expressive musical movement refer to locomotion (Friberg *et al.*, 2000), general embodied movement and gestures (Gritten and King, 2006 & 2011), or metaphorical listening (Budd, 2003; Scruton, 2004; Zangwill, 2010). However, analytical models seem almost exclusively to rely on the measurement of beat distance in order first and foremost to establish questions concerning expressive musical motion.

In his seminal work on musical motion, Truslit (see Repp, 1992) speaks of "melodic motion", implying non-rhythmic motion, rather than the specific interval and rhythmic proportions in a tune or motif. "Melodic motion" is for him the expressive shaping of "intensity" and "duration" of the notes that are constitutive of a melody. Additionally, in his attempt to transcend rhythmic motion, Truslit correlates "inner" motion that has the potential to engage the "whole person" with "organic", i.e., irregular, timing. Truslit's position, hence, may be seen as seeking distance from straightforward physiological models of direct rhythmic "entrainment" (see Trost & Vuilleumier 2013 and Trost *et al.,* 2017).

In the introductory remarks to his translation and synopsis of Truslit's *"Gestaltung and Bewegung in der Musik"* from 1938, Repp (1992, p. 265), while noting Truslit's "breadth", "depth", "wisdom" and "relevance", also points out a "lack of methodological sophistication" in the works of "old authors", deeming them largely 'speculative and subjective". However, one can question Repp's own method of precise acoustic measurement and quantification of musical timing as a reliable pathway to musical motion in the full aesthetic sense that Truslit tried to capture with his framework of motion loops. Indeed, if musical motion is conceptualised one-dimensionally, in terms of a time arrow that sound is supposed to inform exclusively through pulse, metre and rhythm, one abandons features of melodic contour, harmonic tension, texture and dynamic shaping that may ultimately "justify" musical timing. For instance, by simply mapping and differentiating relative distances among (melodic) pitch levels in a musical sequence, one introduces an additional "spatial" variable of contour characteristics that may not only influence perceived tempo, but also the perception of overall motion. While a visual mapping of interval quantities or melodic peaks and valleys is by no means straightforward,[2] musical contour seems to be intrinsically related to vertical spatial connotations (see Romero-Rivas *et al.*, 2018). Here, one may want simply to treat gestural motion as a vector that is separated from musical tempo *per se*, yet such a methodological stance appears prematurely to abandon the possibility of the two vectors being added in the experience of music (see also Moshammer, 2012).

The importance of contour-based or gestural motion can be highlighted by the simple fact that vocal impressions play a significant role in the rehearsing and teaching of instrumental music. Clearly, every instrument is associated with particular timbral qualities and an idiosyncratic acoustic envelope. Yet it is the human voice, as the most original and seemingly precise instrument, that is often best suited to shape and communicate expressive musical intentions. As highlighted and demonstrated in Moshammer (2016), the vocal sketching of musical intentions may make use of diatonically imprecise continuous frequency trails in connecting melodic notes, in order to clarify the gestural dynamic image that should underline a certain discrete tonal sequence. Each (acoustic) instrument has obviously its own capability of sound production that the human voice cannot simply replicate. Importantly, however, being in most cases of traditional music-making bound to discrete tonal systems, the fine-tuning of overall musical motion and expression requires dynamic shaping and also agogic accentuation that in interplay with musical timing generate detailed music expression, the meaning of which calls for active perceptual inference. For instance, a *cantabile* on the piano always results from a mental projection of a smoothened continuous shaping that cannot be a simple mirror image of acoustic reality, given a piano sound's tonal decay. Such consideration relates back to the "noise versus signal" theme that Zimerman seems to allude to in the interview excerpt quoted at the beginning of this Introduction. If music is not sound, and given its widely acknowledged significance, it must produce sense and meaning in allowing the perceiver to reach beyond the data points and values that each note imports into the experience of music.

This article departs from the hypothesis that particularly musical motion informs music's acoustically unexpressed yet significant hidden layer. Musical sound *sui generis* can only afford the perception of what lies beyond and within the notes, yet does not aesthetically instantiate music in a strictly isomorphic sense. As a mainly continuous phenomenon that is superimposed on an otherwise often discrete tonal structure, a competent mental representation of musical motion appears to necessitate forms of active inference that involve, "bottom up", interpolation and smoothing of, as well as, "top-down", abstraction from the acoustic data that music presents to the ear. This article is, however, not primarily interested in concrete theories of perception, such as an evaluation of the predictive coding hypothesis that at least at first glance may be evaluated as resonating with some of this article's ideas (see Heilbron & Chait,

[2] As Robert O. Gjerdingen observes, motion that take into account specific interval relationships faces principal obstacles: "Were a motion-tracking system limited to a uniform speed in traversing the pitch axis, the interval of a perfect fourth would take five times as long to track as the interval of a semitone. Yet musical practice and perception show no strong evidence of such distinctions in linear processes. On the contrary, the very notion 'scale' or 'arpeggio' assumes isochronous performance as the default case, suggesting that a motion-tracking system must be capable of traversing unequal intervals in approximately equal times" (Gjerdingen 1994, p. 347).

2018, for context). On the contrary, the main purpose of the following considerations is a shift "back" from possible *explanantia* regarding musical expression and its perception to a scrutiny of what seems to be an often uncritically preconceived level of the *explananda* regarding the phenomenon of musical motion. The question this article is concerned with is therefore more a proto-task than a real building block in the study of music's expressivity, interrogating what one actually should attempt to grasp in the communication of musical motion, other than modes of a basic musical rubato.

In what follows, this article chooses a rather idiosyncratic tool of discovery, a music visualiser that is able to separate signal from noise at different degrees of resolution, hence bringing about emergent motion properties that curiously inform music's expressive layer. An initial brief comparative discussion of the so-called *AudioVisualizer* that serves this article is followed by case studies of all nine pieces in Robert Schumann's poetic cycle *Forest Scenes*, which underpin the outline of a summarising basic framework of audiovisual cross-modal qualities in relation to music's dynamism. A closer look at *AudioVisualizer's* functioning, particularly in the light of fine-tuned audiovisual synchrony and the adjustment of data resolution in its interplay with emergent qualities, concludes the article.

2 *AudioVisualizer's* competence and performance in light of analogue abstract animation

It was particularly Oskar Fischinger, who at the beginning of the last century experimented with abstract movies that appear as "visual music", an initiative that originated from the context of abstract art and non-narrative movie making more generally (see Kershaw, 1982, for a highly informative dissertation on this subject matter). With modern technology, Fischinger's vision can be powerfully revived. One direction of such a revival runs counter to the historical development, namely Fischinger's influence on the 1940 Walt Disney production *Fantasia*, where music was put on the screen with rich colours and movements. The original intention of "visual music", or abstract music animation, however, appeared to be the search for the essence of motion and rhythm that both music and moving image can share, for which abstraction and information reduction are guiding principles.

Abstract music animation is at present a rather muted phenomenon. Stephen Malinowski's *Music Animation Machine*[3] that renders piano-roll-type moving scores in creative layouts, due to its prominent presence on the YouTube platform, is probably the most popular source of public abstract audiovisual experience. Within the music research community, a seminal point of reference is the "performance worm" that Jörg Langner and Werner Goebl (2003) suggested as an entity that travels through a basic tempo-loudness space. Note, however, that both of the aforementioned animation modes amount fundamentally to not more than the visualisation of digital data, without generating any genuine cross-modal qualities that both music and moving image share. Obviously, the most popular form of creating aesthetic visual correspondence with music is dance. In a more idiosyncratic approach to analogue music tracking, Manfred Clynes experimented with a sentograph (see Clynes, 1980), a paradigmatic simulation tool that generates curves from the pressure and horizontal movement of a finger on an interface during listening to music, hence translating music directly into the sensation of proto-emotions.

More generally, in music research, performance gestures have been studied in order to elicit the gestural and expressive qualities of "embodied" music experience. However, here one has to note that both music notation and the development of new musical instruments have unleashed their own "combinatorial" space of expressive possibilities that surpass biological constraints on human physiology. Indeed, art and music have constantly been pushing the expressive boundaries of the human body itself, such as in ballet or operatic singing. As aforementioned, the human voice is a powerful tool for the "sketching" of musical intentions even where instrumental music is concerned. However, performer/instrument systems can generate sonic events that a voice, singly or in conjunction with others, cannot render. To name only two conventional examples from the classical repertoire, Liszt's *Mephisto Waltzes* generate "diabolic" motion patterns with irregular jumps and abrupt changes in momentum that could hardly be envisaged vocally; while his fourth *Transcendental Étude* for piano, inspired by Victor Hugo's poem *Mazeppa*, loses its musical footing when a horse is acoustically set free, with Mazeppa strapped onto it. The historical emancipation of musical expression from human

[3] See https://www.musanim.com.

physiology through the extension of the performing human body with musical instruments (or, more radically, its replacement by robotic or electronic music-making), together with music notation as a cognitive aid and visualisation tool, exemplify how technology can reorientate the human mind and how expressive imagination may transcend immediate embodiment.

Motor theories of expression perception are inherently related to muscular activity, a claim that the well-known facial-feedback hypothesis paradigmatically stipulates (Dzokoto *et al*., 2014, Neal & Chartrand, 2011), and that research into the relationship between instrumental skills and perceptual abilities further supports (see Hofmann & Goebl, 2014). Yet one can hardly make the formation of aesthetic intentions and skilful music listening solely dependent on the mastery of instruments. After all, conductors often do not play any of the orchestral instruments they direct. More significantly, however, production gestures and physiological processes enable, but are not necessarily isomorphic to, the ultimate sounding of the played acoustic instruments. A simple illustration here is the execution of even scales on the piano with thumb-over and thumb-under techniques, creating a disruptive hand motion that ideally should not be audible in even scale playing.

While original "analogue" music animation can provide a versatile and autonomous communication channel that may inform the aesthetics of music, it is hardly utilised in music research. One notable exception is Nigel Nettheim's (2007) article on armchair conducting that presents computerised animations of motion curves he derives from the work of the German musicologist Gustav Becking. Moshammer (2012) discusses Truslit's motion curves and experiments with hands-on animations as demonstrations of contour-based musical motion, while Moshammer (2016) uses a so-called Pitch-Dynamics Motion Microscope that animates (homophonic) melodies in order to underline the expressive vocal sketching of musical intentions.

This article utilises *AudioVisualizer*,[4] a free and open-source *Chrome* browser extension that uses basic spectral analysis for its dynamic visual mapping of music, which on a basic level creates motion patterns and transformations that appear similar to those of cymatics, i.e., the study of vibration that is made "analogically" visible on a surface such as a membrane (see Animations 12 and 13 for an illustration of this connection) (Ritchie, 2023, pp. 1–11). On first glance, *AudioVisualizer* may seem to amount to nothing else than yet another playful music imaging tool, producing merely flashy, colourful and dense audiovisual correlates that occasionally appear rather random and, in terms of music aesthetics, could in large parts be deemed insignificant. However, the *Chrome* browser extension offers a tableau of user-adjustable parameters that allows for extensive changes to the basic "scenes" that the software initially provides. The developers encourage such adjustments and users can upload newly created scenes onto the program's website. While most user creations adhere to the rather lush visual aesthetics of standard music visualisers, the tool's versatility allows for curious visual reductions that create defined patterns of movement, able to elicit musical micro-expression in audiovisual synchrony.

AudioVisualizer maps a maximum of 512 output spectra lines onto the visual array,[5] employs a colour function, and uses (in the scenes discussed) dots and lines as basic elements that can be adjusted in size and scale. The code calculates the frequency spectrum using the Fast Fourier Transform (FFT) algorithm. The so-called *ButterAudioProcessor* class in the code processes audio data in chunks of 512 samples at a time. The "spectrumJumps" parameter in the process method of the *ButterAudioProcessor* class determines how many bins to jump when calculating the normalised frequency spectrum. If "spectrumJumps" is set to four, then the normalised frequency spectrum will be calculated using every fourth bin. This reduces the number of frequency bins used in the visualisation and can result in smoother, less noisy visualisations. Overall, the frequency bins are established by dividing the magnitude array of the FFT output into several frequency ranges, and the "spectrumJumps" parameter determines how many bins to use when calculating the normalised frequency spectrum.

Two scenes appear particularly efficient in the aforementioned reduction effort, named "Worm" and "DotsAndLines" respectively (see Fig. 1 for an overview of the scene settings). Note that the case studies in this article disregard *AudioVisualizer*'s colouring, with all animations being adjusted to black and white. Apart from the introductory first animation, all subsequent movies are presented in negative mode with a white background, in order to sharpen the structural outline of the visual motion.

[4] See https://chrome.google.com/webstore/detail/audiovisualizer/bojhikphaecldnbdekplmadjkflgbkfh?hl=en.

[5] The full code of *AudioVisualizer* can be found here: https://github.com/afreakk/ChromeAudioVisualizerExtension.

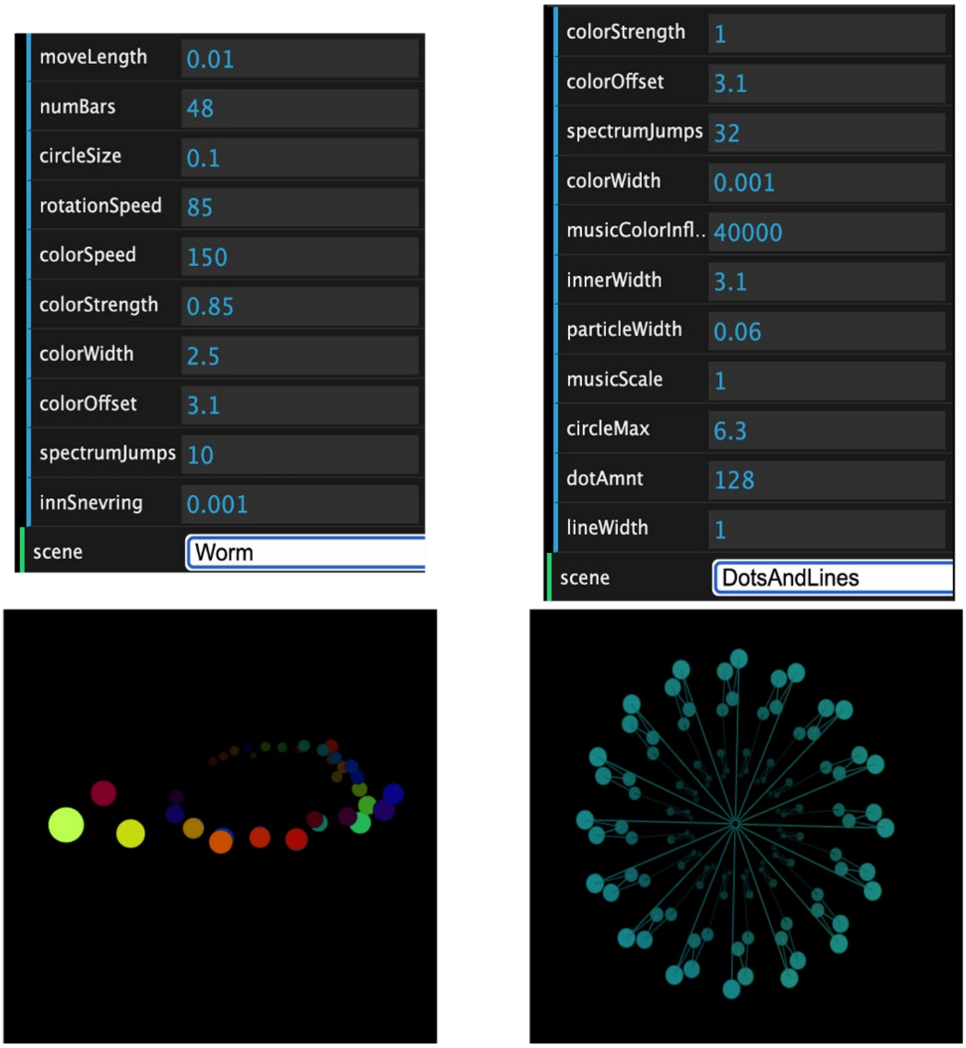

Figure 1: *AudioVisualizer*’s default scene settings for “Worm” and “DotsAndLines” in addition to screenshots of their basic design layout.

The “Worm” scene, which is a kind of moving galaxy of dots, adds rotation speed as an additional aspect. It first calculates the average amplitude of the audio data. The code then divides the average amplitude by the maximum amplitude. It then multiplies the result by the “rotationSpeed” property of the “sceneWormSettings” object. The “rotationSpeed” property is a number that determines the speed of the rotation. The rotation speed is calculated in this way so that the worm-like shape rotates faster when the audio data is louder. In the “DotsAndLines” scene’s initial setting, the dots are evenly spaced along a circle that is centred on the canvas. The average size of the dots is determined by the “particleWidth” property of the “DotsAndLinesSettings” object. Strictly speaking, amplitude only determines changes in dot size, while frequency is responsible for dot positioning. However, because amplitude is related to overtone richness and the visualiser needs to accommodate diverse circle sizes that are loudness-dependent, the resulting imagery captures music’s dynamism quite directly. Furthermore, the so-called “barWidth” parameter influences the interplay between frequency and amplitude by controlling the spacing of the dots. It is a function of the chosen number of dots and the so-called “circleMax” setting, which reads as follows:

barWidth = (xs.circleMax/xs.dotAmnt)/2.

Together with an adjustment of spectrum resolution, this feature significantly influences the visual output of the “DotsAndLines” scene in potentially presenting only a sector of the spiralled array of dots that the original scene setting determines, as well as in producing a particular dot sequencing within such a segment. Here, the visual result is difficult to predict exactly from the onset without hands-on experimentation. The construction mode described can be exploited further by setting the particle width to zero, hence making only the lines that connect the dots visible. Alternatively, in reducing their thickness, one can also let the lines disappear and subsequently create an arrangement of unconnected dots.

What makes this particular music visualisation tool attractive is the interplay between adjustable degrees of spectrum jumps with different amounts of basic elements that emerge in idiosyncratic constellations. Fig, 2 provides an overview of all nine settings used in this article’s case study (see Section 4). The subsequent first animation comprehensively illustrates the series of scenes in Fig, 2 by rendering the continuous apparent upward progression of a so-called Shepard Tone illusion.

1.

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

2.

spectrumJumps	1
colorWidth	0.001
musicColorInfl..	10000000000
innerWidth	1
particleWidth	0.1
musicScale	0.7
circleMax	40
dotAmnt	299
lineWidth	3

3.

spectrumJumps	3
colorWidth	0.001
musicColorInfl..	10000000000
innerWidth	1
particleWidth	0
musicScale	1
circleMax	40
dotAmnt	50
lineWidth	1

4.

spectrumJumps	1
colorWidth	0.001
musicColorInfl..	10000000000
innerWidth	2
particleWidth	0.01
musicScale	0.7
circleMax	6
dotAmnt	9
lineWidth	2

5.

spectrumJumps	1
colorWidth	0
musicColorInfl..	10000000000
innerWidth	3.1
particleWidth	0.2
musicScale	0.6
circleMax	6
dotAmnt	50
lineWidth	0

6.

spectrumJumps	8
colorWidth	0.001
musicColorInfl..	10000000000
innerWidth	0.5
particleWidth	0.00001
musicScale	0.8
circleMax	40
dotAmnt	82
lineWidth	3

7.

spectrumJumps	6
colorWidth	0.001
musicColorInfl..	10000000000
innerWidth	1
particleWidth	0.1
musicScale	0.7
circleMax	40
dotAmnt	82
lineWidth	3

8.

spectrumJumps	2
colorWidth	0.001
musicColorInfl..	10000000000
innerWidth	1
particleWidth	0.1
musicScale	5
circleMax	40
dotAmnt	82
lineWidth	3

9.

spectrumJumps	3
colorWidth	0.001
musicColorInfl..	10000000000
innerWidth	1
particleWidth	0
musicScale	1
circleMax	40
dotAmnt	50
lineWidth	1

Figure 2: Summary of all nine scene settings for the animations of Robert Schumann’s *Forest Scenes* in Section 4. The first table refers to *AudioVisualizer’s* “Worm scene”, the remaining eight to “DotsAndLines”. The numbers correspond to the pieces in Schumann’s cycle.

Animation 1: An illustration of all nine *AudioVisualizer* scene adjustments with the Shepard Tone illusion (click link for animation).

In summary, the functioning of *AudioVisualizer* exemplifies the principle that a higher degree of visual abstraction – i.e., a greater number of spectrum jumps and a smaller number of basic visual elements – does not necessarily lead to reduced expressivity. On the contrary, the "less is more" principle that featured prominently in the Introduction can lead to differentiated expressive properties through the emergence of clearly defined *Gestalts*, for which audiovisual synchronous movement is particularly instructive.

3 Methodological Remarks

The following case studies with *AudioVisualizer* render each of Schumann's pieces in his *Forest Scenes* cycle and demonstrate how a rather direct mapping of acoustic information can lead to versatile emergent properties when musical parameters are synthesised with different degrees of abstraction and resolution. Schumann's *Forest Scenes* paradigmatically exemplify the composer's kaleidoscopic romanticism and expressive plasticity, making it a good candidate for an introductory case study that aims to promote what may on first glance look like a rather standard music visualisation tool to a serious source for music analysis, research and education. Schumann is well-known for his imaginative character pieces that, within the confines of a small form, are able to portray highly individualistic and differentiated musical poetry. More specifically, however, in order to demonstrate the musicological and aesthetic merits of a tool like *AudioVisualizer* successfully, one must look for music that, on the one hand, is dynamically highly differentiated and that, on the other, however, often leaves space between its notes. Dynamic plasticity and textural sparsity seem indeed to be directly correlated with heightened demand for expressive interpretation and active mental projection, for which abstract music animation can offer diverse visual blueprints.

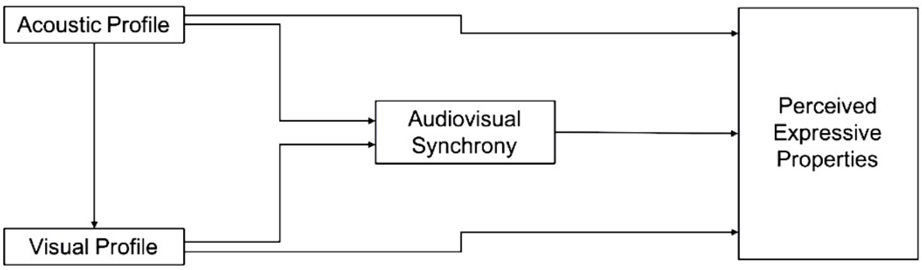

Figure 3: A summary of the perceptual constellation of audiovisual synchrony through automated music visualisation.

With reference to Fig. 3, it can be asked whether audiovisual information adds anything new to the experience of music or, even more fundamentally, whether automated abstract music visualisation can generate relevant expressive visual content that one could not discover in music alone. Alternatively, one may interrogate whether audiovisual experience could actually detract from genuine music listening. While these fundamental questions must ultimately also become informed by empirical perception studies, they do implicate semiotic principles that, within a descriptive methodological research paradigm, can offer valuable insights into questions of music theory and analysis, especially, as stipulated in this article, regarding the notion of musical motion.

In order to get a grip on the aforementioned, it is important first to appreciate the abstract nature of the animations under discussion in this article. Obviously, visual content can add to the interpretation of music (see Platz & Kopiez, 2012). In a scene from *The Errand Boy* (1961), Jerry Lewis performs a pantomime to Count Basie's "Blues in Hoss' Flat". The name of this movie segment is *The Chairman of the Board*, in which, through Jerry Lewis, the sound of a Jazz Big Band becomes the "voice" of a first strict, then furious and finally amused executive chairing a boardroom's empty table. It is remarkable that this scene not only conveys emotions, but a sense of *what* the music in this setting could possibly "say". Musical phrasing, gestures and tension are so clearly marked and so powerfully embodied in Jerry Lewis' comedic pantomime that the ostensibly most important linguistic component, referential semantics, emerges in music without actually being pronounced. While such visual and narrative imposition upon music is not available to abstract animation, what it can achieve is what could be called audiovisual priming of musical motion and gesturality.

As highlighted in the Introduction, the automated abstract visualisation of music introduces a semiotic layer that not only turns what is often discrete score structure into continuous motion, but additionally transcends acoustic reality through processes of data interpolation, smoothing and abstraction. Since this new semiotic dimension is directly derived from acoustic processing, one can formulate the hypothesis that what can actually be seen in automatically generated audiovisual motion formations may correlate with the internal mental processing of musical meaning. Obviously, music's encoding of expressive qualities is by no means limited to motion and movement, yet what "visual music" can offer is the "bracketing" of one crucial component that factors into the constitution of music's signification. It does this by offering novel blueprints for what Charles Sanders Pierce, in the framework of his triadic semiotics, called "interpretants" (see Short, 1996, for context), i.e., the (mental) mediation between physical sign (sound) and its reference (for instance, expressed qualities). Abstract animation offers here idiosyncratic visual proxies for possible interpretants that bridge the gap between sound and meaning. Notably, animation is able to generate interpretative visual imprints of music that are as accessible as bodily motion, such as in dance and performance gestures, yet they are often more versatile and operate with a higher degree of differentiation. In particular, audiovisual motion can pinpoint the integration and synthesis of musical parameters, such as the incorporation of dynamic trajectories into general tempo characteristics and rhythmic patterns. Furthermore, in bridging gaps between notes, visual music tracking gives rise to minuscule motion transpositions that may signify micro-gestures, but also feed the establishment of overall textural qualities. Finally, audiovisual imagery may discover motion universals that go beyond a particular sensory channel, providing a solid foundation for aesthetic metaphorical thinking, ranging from motion that looks like expansive physical actions to more subjective gestural outlines such as the shrugging of one's shoulders.

Before taking a deeper look into the competence and performance of *AudioVisualizer,* the following remarks address the methodological status of the descriptive framework (see Tab. 1 below) that builds on the brief case studies in the following section. It appears important for a methodological evaluation of the subsequent selective descriptions of audiovisual content that if music instils perceptual learning and training that builds competence in grasping its structural and emotional significance, one must acknowledge a normative "top-down" component of mastery and excellence in the constitution of what a particular kind of music actually is. For instance, it usually is sufficient in order to get a correct idea about the pronunciation of, say, a French expression, to consult with a single native speaker of French. While the expert/non-expert distinction in music does not exactly mirror the relationship between native and non-native speaker competence in natural languages, the scientific study of music must incorporate the fact that a schooled or cultured aesthetic sense creates a musical world that would simply not "exist" without such faculty. As much as natural languages and their dialects die when their speakers disappear, forms of (art) music can get lost if the ability for their production and perception vanishes, independently of whether their sound is still preserved on recordings. This very fact seems to complicate further the empirical study of musical meaning and expression: it is an empirically accessible fact itself that first-rate performers of music to this day regularly demonstrate a level of expressive understanding in masterclasses that the general audience may lack. Such an "upper band" of aesthetic excellence and differentiation can, however, be complemented by a "lower band" of assumed competencies that must appear self-spoken in the sense of an "a priori of communication" (see Apel, 1972). This thought recalls principles of hermeneutic understanding and questions regarding the methodological status of the humanities or, for that matter, the sciences that deal with the human mind and its products.

One may suggest that the labelling of Schumann's friendly landscape in the light of Animation 6 below as "fluffy", or, maybe more strikingly, the identification and appreciation of the highlighted micro-gesture in *Lonely Flowers* in Section 4.3 can only be made valid through controlled empirical perception studies with diverse participants. However, the scientific study of cognitive and psychological constraints on feature detection cannot have ultimate bearing on the question of whether certain aesthetic features exist or are ready for discovery. A tool like *AudioVisualizer* serves a more preliminary "archaeological function" here, by making certain expressive qualities of music visually accessible and thus ready for a rather straightforward comparative description and discussion in common natural language. Such a process can hardly be abandoned in preparing a phenomenon for empirical study or, if deemed helpful, in exactly measuring its underlying parameters. What this article basically seeks to underline is that a complex and still inconclusively studied phenomenon as musical motion and expression, particularly in cultural products that transcend basic forms of physiological entrainment, warrants effort in “pre-theoretically” establishing what the *explanandum* under discussion actually might be. Ironically, it is a simple automated algorithm that helps in the current context with this very task.

4 Dynamic Images of Robert Schumann’s *Forest Scenes*[6]

4.1 *Eintritt (Entry)*

The first piece of Schumann’s cycle marks a joyful and lively entry into the forest that the piece’s rhythm accentuates. On top of an exhilarating chord-based rhythm, Schumann places a tune reminiscent of light-hearted singing. The shift from rhythmic motion to melodic deliberation, as shown in Fig. 4 with reference to the piece’s beginning section, coincides in the corresponding visualisation profile (see Animation 2) with a transition from a steadily spinning structure to one that experiences dynamic pushes. This motif of altering momentum gain and momentum decay stretches over the whole animation of the piece, creating an organic motion layer that absorbs rhythm and pulse and that would emerge in any piece that exhibits a dominant dynamic contour. The circles that form the two arms of the spinning structure create a second dimension of dynamic movement, capturing both textural and articulation-related motion details of the performance that are based on amplitude variability as a genuine motion parameter.

Figure 4: Score excerpt of *Entry* from Robert Schumann’s *Forest Scenes*, op. 82,[7] with highlighted bars 3–8 (see Animation 2).

Animation 2: An *AudioVisualizer* rendering of *Entry* from Robert Schumann’s *Forest Scenes*, op. 82 (see Fig. 2 for animation settings).

4.2 *Jäger auf der Lauer (Hunter on the Lookout)*

Hunter on the Lookout is a short musical poem of suspense, tension, nervousness and, ultimately, success. Together with the cycle’s eighth piece, entitled *Hunting Song*, with which it shares a common theme, the second piece in Schumann’s cycle poses a challenge to automated spectrum-based music visualisation that focuses on the discovery of emergent expressive qualities. In pieces with fewer dynamic fluctuations, especially against the backdrop of extensive *forte* passages, pitch perception predominantly carries expressive shaping, without being indirectly accessible in the visualisation of corresponding dynamics, which essentially characterises the animation modus that this article discusses. In order to mitigate this challenge, Animation 3 employs a relatively high resolution with a low number of spectrum jumps and a high number of circles as data points, which leads to a busy pattern that is susceptible to minor

[6] The recording used in the animations is by Sviatoslav Richter, dated 1956 and published by *Deutsche Grammophon*.
[7] Source: *Waldscenen. Neun Klavierstücke*, op. 82. Breitkopf & Härtel, Leipzig 1882 (edited by Clara Schumann). All score figures in this article derive from the same source.

"dynamic" fluctuations. From this mode of representation emerges in the given case, quite appropriately one could say, a trembling texture that expresses suspense and nervousness. Fig. 5 highlights a simple motif of quivering quaver triplets at the piece's ending that characterises the storyline right from the point where it follows the piece's opening gesture of being "on the lookout". What makes the indicated passage particularly noteworthy is its conclusive shift to D major, offering the bright colour of success, with the said motif exhibiting a smoother visual fluctuation in the animation, one that in audiovisual synchrony hints at the sensation of restrained excitement rather than anxiety.

Figure 5: Score excerpt of *Hunter on the lookout* from Robert Schumann's *Forest Scenes*, op. 82, with highlighted bars 33 and 34 (see Animation 3).

Animation 3: An *AudioVisualizer* rendering of *Hunter on the lookout* from Robert Schumann's *Forest Scenes*, op. 82 (see Fig. 2 for animation settings).

4.3 *Einsame Blumen (Lonely Flowers)*

While *Lonely Flowers* definitely delivers touches of melancholy, it also sets a friendly tone of simplicity, unpretentious beauty and content. In Richter's rendering, the piece's moderate pace gives importance to every quaver of its simple melody. The corresponding animation profile operates with a higher degree of abstraction than the rendering in Animation 3. However, the lower number of dots set in Animation 4 increases its gestural profile, creating defined shifts from one melodic note to another in capturing dynamic profiling and articulation. Within this array of permanent gestural alterations in moving from one quaver to the other, there are occasional broader *Gestalts* emerging, indicating clear expressive performance intentions in directing the groups of four quavers that shape the piece. For instance, bar 14 (see Fig. 6) appears in Animation 4 as a subtle uplift in the transition from the first quaver to the second, with the group ending in a soft retreat, indicating a curious interplay between animation settings and the emergence of micro-gestures in correspondence with dynamic shaping and timing.

Figure 6: Score excerpt of *Lonely Flowers* from Robert Schumann's *Forest Scenes*, op. 82, with highlighted bar 14 (see Animation 4).

Animation 4: An *AudioVisualizer* rendering of *Lonely Flowers* from Robert Schumann's *Forest Scenes*, op. 82 (see Fig. 2 for animation settings).

4.4 *Verrufene Stelle (Haunted Place)*

Animation 5 shapes the mystery of Schumann's *Haunted Place* with a simple formation of nine connected dots that however do derive from the full range of spectrum analysis. This visualization modus leads to a versatile geometric body with the ability to support the finesse of the composer's gestures of suspense. Indeed, one can even personify the abstract figure that, while moving along with the music in audiovisual synchronization, expresses cautious fear in nevertheless remaining faceless. Fig. 7 points to a particularly paradigmatic constellation, one where a gestural transition from high to low register creates a rather awkward audiovisual "body language" of directional shifts, supporting the piece's expressive program with utmost efficiency in evoking emotional allusions.

Figure 7: Score excerpt of *Haunted Place* from Robert Schumann's *Forest Scenes*, op. 82, with highlighted bars 9–13 (see Animation 5).

Animation 5: An *AudioVisualizer* rendering of *Haunted Place* from Robert Schumann's *Forest Scenes*, op. 82 (see Fig. 2 for animation settings).

4.5 *Freundliche Landschaft (Friendly Landscape)*

What is a friendly landscape in Schumann's fifth piece of his *Forest Scenes* becomes a light and airy image in Animation 6. The high number of larger dots creates a garland that alludes to the motion of connected balloons due to the lightness of their movement. Fig. 8 paradigmatically highlights first a passage that, throughout the piece, and in different variations, appears as the conclusion to the busy "airborne" phrases that characterise the music. Throughout the piece, this resolution in a minim-crotchet pattern momentarily slows down the motion, with each of the transitions exhibiting diverse expressive details that in the animation results in various exemplifications of depreciating inner tension. At the end of the piece, however, as additionally highlighted in Fig. 8, the lively melody is pushed up, which echoes a similar motion that occurs at the end of the piece's first part, only to lead ultimately to a heightened gesture of final joy with a $b\flat^2$–g^2–$e\natural^2$ motif, especially due to its repetition. This final denial of withdrawal counters expectation, for which the piece up to this moment has laid the groundwork with the regularity of its more restrained phrase-endings.

Figure 8: Score excerpt of *Friendly Landscape* from Robert Schumann's *Forest Scenes*, op. 82, with highlighted bars 43–46 and 49–52 (see Animation 6).

Animation 6: An *AudioVisualizer* rendering of *Friendly Landscape* from Robert Schumann's *Forest Scenes*, op. 82 (see Fig. 2 for animation settings).

4.6 *Herberge (Wayside Inn)*

Taking a rest, recharging one's batteries, and filling one's stomach are usually rewarding events. Schumann's *Wayside Inn* tries to capture the positive spirit of such a break, rendering it in a joyful tune that, especially with its upwards jumps into a dotted rhythm, energises body and soul, before the piece finally ends in slumberous content. Animation 7 captures the piece's kinematics with bouncing arms that are arranged in a mirror image, which, in an admittedly speculative interpretation, could be read as playful interpersonal interaction. However, the visualisation also gives insights into the motion of articulation. Fig. 9 shows passages of soft *portato*-like touches that let the animation's arms float in space, with the second highlighted section showing a clear downbeat and accent on the F minor chord that, in the animation, coincides with the occurrence of a blue background for easier identification.

Figure 9: Score excerpt of *Wayside Inn* from Robert Schumann's *Forest Scenes*, op. 82, with highlighted bar 39 and bars 43 and 44 (see Animation 7).

Animation 7: An *AudioVisualizer* rendering of *Wayside Inn* from Robert Schumann's *Forest Scenes*, op. 82 (see Fig. 2 for animation settings).

4.7 *Vogel als Prophet (Bird as Prophet)*

Maybe the most famous piece in Schumann's work, *Bird as Prophet* sets a tune from another world, delivering estranged, mysterious and seductive sounds. The piece is framed in an A–B–A form and the phrases in section A end with a kind of musical question mark, reaching out to an unknown territory. Animation 8 shows a skeleton-like pattern with pronounced geometric transitions that result from a relatively high level of spectrum abstraction and the rather subdued, yet dynamically differentiated, soundscape of the piece. This modus shapes the Bird's evocations with subtle mechanical micro-eruptions, as if they were curious calls of the future. Fig. 10 highlights two rhythmic motifs that, in the visualisation, are particularly striking in pointing upwards due to the bright sounding of the a^3 that finishes the pattern (while in other motivic variations of the same rhythm the gesture closes inwards).

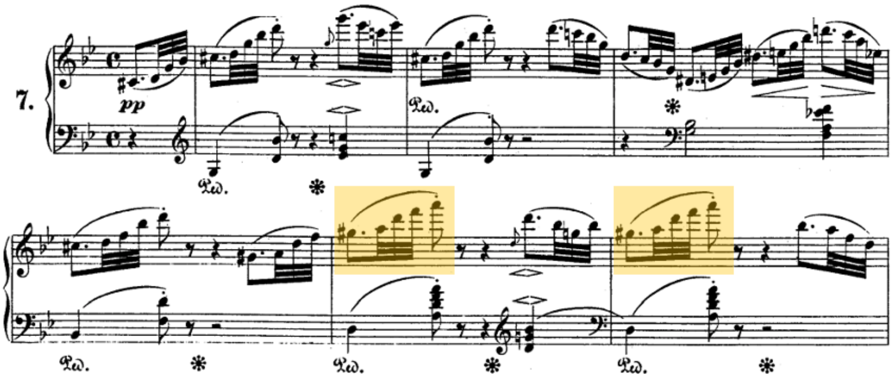

Figure 10: Score excerpt of *Bird as Prophet* from Robert Schumann's *Forest Scenes*, op. 82, with highlighted bars 5 and 6 (see Animation 8).

Animation 8: An *AudioVisualizer* rendering of *Bird as Prophet* from Robert Schumann's *Forest Scenes*, op. 82 (see Fig. 2 for animation settings).

4.8 *Jagdlied (Hunting Song)*

As mentioned, the eighth scene is the second occurrence of the hunting motif in Schumann's cycle. *Hunting Song* is the most dynamic and powerful piece in the collection, with sharp accents that lend themselves to visual correspondence. Apart from these signals of joy, strength and power that, especially in the piece's B section, almost shoot like arrows, Animation 9 generates a vibrant dynamism that results from the same visualisation modus that illustrates the prophetic bird in Animation 8. However, a higher spectral resolution, upscaling and an additional zooming-in on the centre of the "skeleton" creates enhanced visual plasticity, drawing the viewer more into the centre of the motion. Fig. 11 emphasises the rhythmic transition from the piece's first part to its B section, which creates the effect of absorption that transitions from large-scale high density to medium-scale plasticity, adding texture and motion characteristics to what seems merely to be a simple rhythm in *diminuendo*.

Figure 11: Score excerpt of *Hunting Song* from Robert Schumann's *Forest Scenes*, op. 82, with highlighted measure 16 (see Animation 9).

Animation 9: An *AudioVisualizer* rendering of *Hunting Song* from Robert Schumann's *Forest Scenes*, op. 82 (see Fig. 2 for animation settings).

4.9 *Abschied (Farewell)*

The final piece of Schumann's *Forest Scenes* is a peaceful farewell with both touches of melancholy and brightness. Animation 10 captures this ambivalent atmosphere with a shimmering, flower-like display, exhibiting an organically altering inner structure with the contours of the flower's petals becoming gestural. Fig. 12 points to simple downward gestures between c^2 and $e\flat^1$, as well as c^2 and f^1, that the animation exemplifies as active inward motion, again highlighting movement between the notes, instead of being limited to rhythm and pulse.

Figure 12: Score excerpt of *Farewell* from Robert Schumann's *Forest Scenes*, op. 82, with highlighted bars 3 and 4 as well as 7 and 8 (see Animation 10).

Animation 10: An *AudioVisualizer* rendering of *Farewell* from Robert Schumann's *Forest Scenes*, op. 82 (see Fig. 2 for animation settings).

4.10 A framework for integrated musical motion and its expressive derivatives

As demonstrated above, music's animacy and movement should not be reduced to questions of tempo and timing alone. In particular, Animation 2, by utilising *AudioVisualizer*'s rotation feature, underlines this point. Animations 2–10, each in their own way, demonstrate that musical motion is also intimately related to amplitude, or, in psychoacoustic terms, to loudness, articulation and even timbre. Motion is, however, not the only cross-modal quality the series of animations in this article helps discover and make concrete. As a medium of continuously shaped dynamism, music is able to evoke a powerful register of cross-modal associations (see Tab. 1 for a synopsis).

Piece (animation number)	Cross-modal signature	Specific features
Entry (2)	Dynamic motion	Motion pushes and pulls that emancipate from yet are integrated in musical tempo
Hunter on the Lookout (3)	Exemplification	Trembling
Lonely Flowers (4)	Micro-gestures	Idiosyncratic continuous dynamic contours that transcend discrete tonal structure
Haunted Place (5)	Animacy and personification	"Awkward" cautious motion and expression
Friendly Landscape (6)	Expression ("synaesthesia")	Airiness, lightness, fluffiness, softness
Wayside Inn (7)	Articulation	Impulsive and soft accentuation
Bird as Prophet (8)	Velocity and momentum	Complex kinematic motion
Hunting Song (9)	Rhythmic texture	Energetic outburst and abrupt motion alterations
Farewell (10)	Spectral texture	Continuously floating, shimmering and shining movement

Table 1: Cross-modal properties and selected expressive characteristics in Animations 2–10.

Nelson Goodman, in his seminal recalibration of symbol theory (1968, pp. 52–57 and pp. 85–95), introduces exemplification and expression as genuine modes of reference that play a significant role in the functioning of art that traditionally have often been rendered "formalistic" (see, for further analysis, Moshammer, 2017, pp. 274–276 and Moshammer & Ekamp, 2018, pp. 8–10). While music may rarely be denotative, i.e., refer to entities and qualities it does not instantiate, it can direct our attention to certain features it possesses, literally or metaphorically. The trembling in Animation 3, during the hunter's tension of being on the lookout, and the *Friendly Landscape*'s airy feel and almost tactile "synaesthetic" fluffiness, draw attention to qualities that are not reserved to music alone and that, if metaphorical, sound can only indirectly express. Exemplification and expression can evoke specific qualities, such as weight of articulation (Animation 7) and directed micro-expression in the form of minuscule gestures (Animation 4); but they can equally extend to a focus on more global textural features (Animations 9 and 10). Animation 5, rendering the *Haunted Place*, is of particular interest because it obtains the function of an animated quasi-subject, a (virtual) *persona* (Lidov, 1999, p. 219), which alludes to music as an artificial agent that attracts empathy, i.e., an invitation to "moving along". In contrast, the kinematic motion that emerges in the chosen rendering of the prophetic bird alludes to an inanimate mechanism (Animation 8; see Moshammer 2012 for a more detailed discussion of a differentiation between subjective "animated" and objective "physical" motion).

Both sound and moving image exhibit the enlisted expressive qualities in Tab. 1 quite literally. However, expressive features that are metaphorical in Goodman's sense, i.e., referencing non-visual and possibly even non-acoustic properties, appear to necessitate genuine audiovisual experience. It is the music that makes the prophetic bird mysterious, and Schumann's *Farewell* could hardly be associated with melancholic content through the associated animation alone, without sounds of reflective reminiscence. Musical tension alone is intrinsically related to the harmonic relationships (see Farbood 2012) that crucially underpin music's emotional signature. Hence, audiovisual motion isolates only one of many contributing variables of music's expressivity. Yet in establishing a highly differentiated, intersubjectively accessible, concrete motion image that otherwise is hidden in processes of mental presentation, abstract animation has the potential to guide the overall comprehension of musical expression in terms of a novel semiotic layer.

5 Audio-visual synchrony, abstraction and emergence: A closer look at *AudioVisualizer's* performance characteristics and analytical potential

Micro-expression can emerge from reduced acoustic fidelity, or, in other words, by separating signal from noise in a creative process of construction. In selecting a particular visualisation mode that exhibits continuous motion, *AudioVisualizer* is able to produce a wide range of detailed imprints that can guide the audiovisual experience of music. Such visual guidance and determination of the acoustic experience does not seem to be overruled by the insight that auditory rhythm appears more accurate than vision in the diachronic structuring of movement (Repp & Penel, 2004). Note that audiovisual synchronisation is a paradigmatic case of a wide range of multi-sensory integration, such as speech-lip (Lewkowicz, 2010; Vroomen & Stekelenburg, 2011) and speech-gesture (Habets *et al.*, 2011), for instance. In relation to music perception, however, respective investigations have mainly made contributions to the correspondence of audio and visual rhythm (Gomez-Ramirez *et al.*, 2011), as well as to the coordination of audio rhythm with biological motion (Phillips-Silver & Trainor, 2007). Here, the main focus is on the scrutiny of the perceptually most important "amodal" property, timing, since temporal proximity is necessary for the perceptual binding of multisensory input. Subsequently, the definition of "horizons of simultaneity" in diverse AV contexts becomes a major target (see Keetels & Vroomen, 2012). Yet the two prevalent research methods – the temporal order judgment task and the simultaneity judgment task (see Love *et al*., 2013) – while providing some interesting psychoacoustic measures, do not immediately suggest any deeper insights into the phenomenology of music. This applies equally to carefully investigated phenomena such as the Schutz-Lipscomb illusion (Schutz & Kubovy, 2009) of visual cues influencing the perception of acoustic duration. Some studies have taken new directions in the testing of audiovisual synchrony under the consideration of diverse musical parameters. Experiments that draw from standard music visualisers (see, for instance, Mossbridge *et al*., 2012) are however, aesthetically rather inconclusive.

In order to demonstrate the aesthetic subtleties and perceptual challenges that are associated with audiovisual priming, Animation 11, by employing an additional scene setting (see Fig. 13), presents a sequence that compares the beginning of Schumann's *Bird as Prophet* with its recurrence after the middle section. Richter's performance of the two passages is obviously slightly different, allowing a comparison of the visualisation of the two passages in terms of synchronisation with both matching and mismatching sound (see Fig. 14 for details).

spectrumJumps	1
colorWidth	0.001
musicColorInfl..	10000000000
innerWidth	3
particleWidth	0
musicScale	1
circleMax	6
dotAmnt	25
lineWidth	0

Figure 13: *AudioVisualizer* profile of the new scene employed in Animation 11. In terms of "spectrum jumps", this scene is a slightly more detailed and, hence, "fluid" version of the visualisation modus that renders the animation of Schumann's *Haunted Place.*

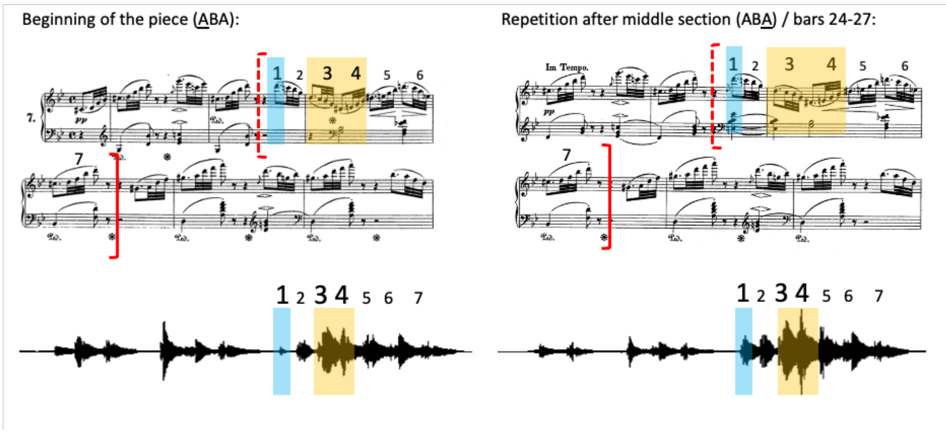

Figure 14: Two almost identical score excerpts from Robert Schumann's *Bird as Prophet* with their associated waveforms deriving from Sviatoslav Richter's performance. The two passages, named simply A and B, are used in Animation 11 in short illustrations of matching and mismatching audiovisual synchrony, especially regarding the highlighted transition from the indicated notes "3" to "4".

Animation 11: A series of side-by-side and overlapping comparisons of matching and mismatching audiovisual synchrony regarding the passages A and B and their respective interpretation by Sviatoslav Richter as indicated in Fig. 14.

The rather minimal acoustic differences between Richter's interpretation of the two almost identical score passages may be difficult to grasp, and they illustrate the phenomenon of "multistability" (Schwartz *et al.*, 2012), i.e., a certain flexibility in perceptual audiovisual matching. Yet, as is especially highlighted in the closing section of Animation 11 that applies the utilised new visualisation modus at half-speed, with automated abstract animation one can make immediate "analogue" sense of the miniscule expressive differences between the two sound events, especially regarding the transition between note "3" and "4", as indicated in Fig. 14.

The two scenes to which Animation 11 applies exhibit significant differences in their motion characteristics, yet both are based on the identical *AudioVisualizer* scene "DotsAndLines". The rather surprising expressive versatility of *AudioVisualizer* derives from the interplay between data abstraction and expressive emergence. In order to provide a clearer image of this analogue process, Animation 12 exemplifies the transition from a rich high-resolution rendering (see Fig. 15 for the respective parameters) to a defined emergent expressivity with a comparative rendering of Schumann's *Bird as Prophet*. In Animation 13, referring back to Animation 5 of

Schumann's *Haunted Place*, this process of abstraction appears even more radical. Additionally, this final animation rotates the original animation 90 degrees clockwise, demonstrating the establishment of a slightly altered character in such simple visual reorientation.

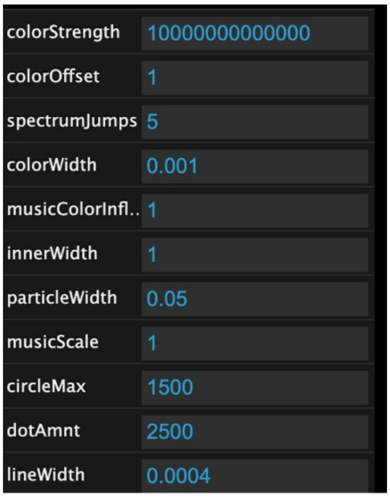

Figure 15: *AudioVisualizer* profile of the high-resolution scene that is used as a template for the comparisons in Animations 12 and 13.

Animation 12: Overlapping comparison of two visualisations of Robert Schumann's *Bird as Prophet* with two individualised scenes from *AudioVisualizer*'s "DotsAndLines" template, one in high resolution, the other in lower resolution.

Animation 13: Side-by-side comparison of two visualisations of Robert Schumann's *Haunted Place* with two individualised scenes from *AudioVisualizer*'s "DotsAndLines" template, one in high resolution, the other in lower resolution.

In summary, while the multitude of emergent motion patterns that this article presents may be difficult to systematise, it is important to acknowledge that they result from a rather simple and straightforward mapping algorithm, with the ability to engage in heterogeneous forms of abstraction and resolution, without at any point losing "analogue" contact to the factual acoustic happening. This insight bridges the gap between hard acoustic data and the versatility of musical imagination, not only in terms of music's global features, but also in relation to its subtlest sonic micro-expression.

6 Conclusion

It is obvious that "visual music" needs sound in order to make complete sense, while sound does not depend on visual cues in order to convey its message. However, music visualisation can function both as a tool of discovery and of priming, thus informing and potentially refining the process of music listening. In addition to this didactic function, studying modes of music visualisation relates to the broader aesthetic question as to why certain types of music, or individual pieces, respond "better" to particular animation styles than others, which is especially meaningful if the animation is automated, i.e., algorithmically derived from acoustics. Finally, abstract animation may stimulate scientific research regarding the functioning of human perception, because each animation profile that this article applies could, for instance, be evaluated as a particular "brain" that synthesises and interprets acoustic information.

More specifically, the considerations in this article may motivate the creation of novel audiovisual stimuli for empirical perception studies. In addition to investigations into the functioning and perceptual boundaries of audiovisual multistability, as illustrated in Animation 11, abstract animation lends itself to the empirical examination of how various modes of visual priming may guide and differentiate an expression-focused evaluation of identical acoustic stimuli. Such a procedure implicates questions as to the perceived expressive and aesthetic adequacy or individual preferences in relation to diverse animation modes. The interplay between audiovisual imaging and the discovery of expressive features may be further studied concerning its possible potential in enhancing the verbal description of musical content in terms of concrete motion formations, which may prove particularly significant in the audiovisual listening of non-musicians.

Finally, since audiovisual synchrony generates a continuous image that is superimposed on what often functions as a discrete tonal structure, one could further employ machine learning and neural networks in order potentially to evaluate motion-based expressive features of music from a new angle. Here, larger data sets of audiovisual renderings may allow for general conclusions as to the stylistic idiosyncrasies of diverse performers and composers, or even for the discovery of more comprehensive characteristics of generic music styles and genres. Such deeper insights into the functioning of musical expression, and the discovery of new and significant patterns of movement require, however, a clearer understanding of how continuous emergent "interpretants" of music can be classified and modelled, particularly in the light of detailed motion formations. Abstract animation provides one possible entry point to such a research agenda, for which this article has attempted to provide a modest stepping stone.

References

Adorno, T. W. (1963). *Der getreue Korrepetitor: Lehrschriften zur musikalischen Praxis*. am Main: S. Fischer Verlag.

Apel, K. O. (1972). The a priori of communication and the foundation of the humanities. *Man and World*, *5*(1), 3–37.

Budd, M. (2003). Musical movement and aesthetic metaphors. *The British Journal of Aesthetics*, *43*(3), 209–223.

Clynes, M. (1980). The communication of emotion: Theory of sentics. In R. Plutchik & H. Kellerman (Eds.), *Theories of emotion* (pp. 271–301). Academic Press.

Dittrich, W. H., Troscianko, T., Lea, S. E., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, *25*(6), 727–738.

Dzokoto, V., Wallace, D. S., Peters, L., & Bentsi-Enchill, E. (2014). Attention to emotion and non-western faces: Revisiting the facial feedback hypothesis. *The Journal of general psychology*, *141*(2), 151–168.

Farbood, M. M. (2012). A parametric, temporal model of musical tension. *Music Perception*, *29*(4), 387–428.

Friberg, A., Sundberg, J., & Frydén, L. (2000). Music from motion: Sound level envelopes of tones expressing human locomotion. *Journal of New Music Research*, *29*(3), 199–210.

Gabrielsson, A. (2003). Music performance research at the millennium. *Psychology of music*, *31*(3), 221–272.

Gjerdingen, R. O., Todd, P., & Griffith, N. (1994). Apparent motion in music? *Music Perception, 11*(4), 335–370.

Goebl, W., Dixon, S., & Schubert, E. (2014). Quantitative methods: Motion analysis, audio analysis, and continuous response techniques. In D. Fabian, R. Timmers & E. Schubert (Eds.), *Expressiveness in music performance: Empirical approaches across styles and cultures* (pp. 221–239). Oxford University Press.

Gomez-Ramirez, M., Kelly, S. P., Molholm, S., Sehatpour, P., Schwartz, T. H., & Foxe, J. J. (2011). Oscillatory sensory selection mechanisms during intersensory attention to rhythmic auditory and visual inputs: a human electrocorticographic investigation. *Journal of Neuroscience*, *31*(50), 18556–18567.

Goodman, N. (1968). *Languages of art: An approach to a theory of symbols*. Bobbs-Merrill.

Gritten, A., & King, E. (Eds.). (2006). *Music and gesture*. Ashgate Publishing, Ltd..

Gritten, A., & King, E. (Eds.). (2011). *New perspectives on music and gesture*. Ashgate Publishing, Ltd..

Habets, B., Kita, S., Shao, Z., Özyurek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. *Journal of cognitive neuroscience*, *23*(8), 1845–1854.

Heilbron, M., & Chait, M. (2018). Great expectations: Is there evidence for predictive coding in auditory cortex? *Neuroscience*, *389*, 54–73.

Hofmann, A., & Goebl, W. (2014). Production and perception of legato, portato, and staccato articulation in saxophone playing. *Frontiers in psychology*, *5*, 690.

Honing, H. (2005). Is there a perception-based alternative to kinematic models of tempo rubato? *Music Perception*, *23*(1), 79–85.

Keetels, M. N., & Vroomen, J. (2012). Perception of synchrony between the senses. In M. M. Murray, & M. T. Wallace (Eds.), *The neural bases of multisensory processes* (pp. 147–177). CRC Press.

Kershaw, D. (1982). *Tape music with absolute animated film: Prehistory and development* (Doctoral dissertation, University of York).

Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of cognitive neuroscience*, *26*(7), 1546–1554.

Langner, J., & Goebl, W. (2003). Visualizing expressive performance in tempo-loudness space. *Computer Music Journal*, *27*(4), 69–83.

Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Developmental psychology*, *46*(1), 66–77.

Lidov, D. (1999). *Elements of semiotics*. Macmillan,

Love, S. A., Petrini, K., Cheng, A., & Pollick, F. E. (2013). A psychophysical investigation of differences between synchrony and temporal order judgments. *PloS one*, *8*(1), e54798.

Moshammer, G. (2012). Contour, motion and gesture in abstract score animation: A first approach. *Journal of Music Research Online*, *3*. https://www.jmro.org.au/index.php/mca2/article/view/61/30.

Moshammer, G. (2016). Seeing Carlos Kleiber's vocalised intentions through the Pitch-Dynamics Mo-

tion Microscope. *Journal of Music Research Online*, *7*. https://www.jmro.org.au/index.php/mca2/article/view/160.
Moshammer, G. (2017). Routes for roots: A mapping shorthand symbolism with reference to Nelson Goodman's hidden Ars Combinatoria. *History and Philosophy of Logic*, *38*(3), 263–281.
Moshammer, G., & Ekamp, B. (2018). Inside-out or outside-in? On freeing aesthetic emotions. *Literature & Aesthetics*, *28*(2), 1–20.
Mossbridge, J., Grabowecky, M., & Suzuki, S. (2012). Seeing the song: Left auditory cortex tracks auditory-visual dynamic congruence. *Journal of Vision*, *12*(9), 614–614.
Neal, D. T., & Chartrand, T. L. (2011). Embodied emotion perception: Amplifying and dampening facial feedback modulates emotion perception accuracy. *Social Psychological and Personality Science*, *2*(6), 673–678.
Nettheim, N. (2007). How world views may be revealed by armchair conducting: composer-specific computer animations. *JMM: The Journal of Music and Meaning*, *5*. http://www.musicandmeaning.net/issues/showArticle.php?artID=5.6.
Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition*, *105*(3), 533–546.
Platz, F., & Kopiez, R. (2012). When the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Perception*, *30*(1), 71–83.
Repp, B. H. (1989). Expressive microstructure in music: A preliminary perceptual assessment of four composers' "Pulses". *Music Perception*, *6*(3), 243–273.
Repp, B. H. (1993). Music as motion: A synopsis of Alexander Truslit's (1938) *Gestaltung und Bewegung in der Musik*. *Psychology of Music*, *21*(1), 48–72.
Repp, B. H. (1995). Expressive timing in Schumann's "Träumerei": An analysis of performances by graduate student pianists. *The Journal of the Acoustical Society of America*, *98*(5), 2413–2427.
Repp, B. H., & Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological research*, *68*, 252–270.
Ritchie, L. (2023). Multisensory music performance with cymatic images. *Music & Science*, *6*. https://journals.sagepub.com/doi/full/10.1177/20592043231159065.
Romero-Rivas, C., Vera-Constán, F., Rodríguez-Cuadrado, S., Puigcerver, L., Fernández-Prieto, I., & Navarra, J. (2018). Seeing music: The perception of melodic "ups and downs" modulates the spatial processing of visual stimuli. *Neuropsychologia*, *117*, 67–74.
Schutz, M., & Kubovy, M. (2009). Causality and cross-modal integration. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1791.
Scruton, R. (2004). Musical movement: a reply to Budd. *The British Journal of Aesthetics*, *44*(2), 184–187.
Schwartz, J. L., Grimault, N., Hupé, J. M., Moore, B. C., & Pressnitzer, D. (2012). Multistability in perception: binding sensory modalities, an overview. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1591), 896–905.
Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge University Press.
Short, T. L. (1996). Interpreting Peirce's interpretant: A response to Lalor, Liszka, and Meyers. *Transactions of the Charles S. Peirce Society*, *32*(4), 488–541.
Todd, N. P. M. (1995). The kinematics of musical expression. *The Journal of the Acoustical Society of America*, *97*(3), 1940–1949.
Trost, W.. & Vuilleumier, P. (2013). Rhythmic entrainment as a mechanism for emotion induction by music: a neurophysiological perspective. In Cochrane, T., Fantini, B., & Scherer, K. R. (Eds.). (2013). *The emotional power of music: Multidisciplinary perspectives on musical arousal, expression, and social control* (pp. 213–225). Oxford University Press.
Trost, W. J., Labbé, C., & Grandjean, D. (2017). Rhythmic entrainment as a musical affect induction mechanism. *Neuropsychologia*, *96*, 96–110.
Vroomen, J., & Stekelenburg, J. J. (2011). Perception of intersensory synchrony in audiovisual speech: Not that special. *Cognition*, *118*(1), 75–83.
Zangwill, N. (2010). Scruton's musical experiences. *Philosophy*, *85*(1), 91–104.