Music, animacy, and rubato: what makes music sound human?

Adrian S. Blust, David J. Baker, Kaitlin Richard, Daniel Shanahan

***Abstract*—**Our understanding of—and preference for—music is dependent upon the perception of human agency. Listeners often speak of how computer-based performances lack the "soul" of a human performer. At the heart of perceived animacy is causality, which in music might be thought of as rubato, and other variations in timing. This study focuses on the role of variations in microtiming on the perceived animacy of a musical performance. Recent work has shown that the perception of visual animacy is likely categorical, rather than gradual. Although a number of studies have examined auditory animacy, there has been very little research done on whether it might be thought of as a dichotomy of alive/not alive, rather than a continuum. The current study examines the specific intricacies of musical animacy, specifically, how microtiming variations of inter-onset intervals contribute to the perception that a piece was human performed. Additionally, this study aims to examine the possible nature of categorical/continuous perception of musical animacy. In Experiment 1: "Rohum", computer sequenced MIDI renditions were manipulated to contain set random fluctuations of inter-onset intervals. In Experiment 2: "Humbot", participants were presented with human performances digitally recorded using MIDI keyboards, and were asked how "alive" each performance sounded using a 7-point Likert scale. Human performances were divided into ten degrees of quantization strength, increasing from raw performance to 100% quantization. Results suggest an optimal level of quantization strength that is correlated with higher perceived animacy, and fixed random fluctuations of IOI are not a good indicator of human performance. This paper discusses the role of external stylistic assumptions on perceived performances, and also takes into account musical sophistication indices and experience.

***Keywords*—**Animacy, performance, microtiming variations, rubato.

I. Introduction

THE perception of causality is paramount to the understanding of our surroundings. It is the difference between a falling branch and a stick being thrown at us, between leaves rustling and someone walking behind us. Research in vision has attempted to identify visual cues that are used by the viewer in order to discern whether particular objects are perceived as animate or inanimate. Some of the earliest work on the perception of animacy and causality was conducted by Heider and Simmel who demonstrated the role of motion in the perception of animacy in shapes [1]. Stewart showed that the perception of animacy needs only a few very simple perceptual cues, and experiments by Premack and his colleagues later showed that certain movements allow viewers to ascribe intentions and motives to abstract geometric figures such as triangles and squares [2]–[4]. Tremoulet & Feldman demonstrated that the perception of causality in visual stimuli might be linked to change in direction [5]. Scholl and Tremoulet refer to this as the "energy violation" hypothesis [6]. Broze codified much of this research into six indicators of an animate object [7]:

1. Animate agents are self-propelling or locomotive, and can move under their own power.
2. Animate agents possess intentionality, and exhibit goal-directed behavior.
3. Animate agents are communicative, and employ signalling systems.
4. Animate agents are sentient, and can experience subjective feelings, percepts, and emotions.
5. Animate agents are intelligent, capable of rational thought.
6. Animate agents can be self-conscious, having metacognitive states about their own consciousness and that of others.

It could be argued, however, that these qualities themselves are less important than the perception that an agent has the capacity to convey them. Recent work by Looser and Wheatley has shown that the perception of visual animacy can be understood in agents that lack any locomotive qualities [8]. The study presented photographs showing a sequence of a gradual transition between inanimate and animate faces (dolls and humans, via image manipulation), and asked participants to identify the point where the face "becomes alive". Interestingly, the perception of animacy seemed to be categorical, rather than continuous. Similarly Wheatley, Milleville, and Martin carried out fMRI scans as participants viewed the sequence of morphed photos, and found a change of inferior temporal activation consistent with the animate/inanimate distinction [9] (for similar results, see [10]). While motion, intentionality, or the ability to communicate were not present in these images, given a convincingly animate face, it could be inferred that the image was that of an animate being and thus demonstrate the potential for all of the

A. S. Blust is a recent graduate of the University of Virginia, Department of Cognitive Science & Music, Charlottesville, VA 22903 USA (phone: 202-380-7099; e-mail: asb8nk@virginia.edu)

D. J. Baker is with the School of Music, Louisiana State University, Baton-Rouge, LA, 70802 USA (phone: 414-736-7948; e-mail: Davidjohnbaker1@gmail.com).

K. Richard is with the School of Music, Louisaina State University, Baton-Rouge, LA 70802 USA, (e-mail: richard.kaitlin@yahoo.com)

D. Shanahan is a professor of Music Theory with the School of Music, Louisiana State University, Baton-Rouge, LA 70802 USA, (phone: 614-940-2560; e-mail: daniel.shanahan@gmail.com)

above, rather than demonstrating the explicit ability to do so.

Just as visual cues can be used to infer the animacy of an object, it would follow that auditory cues can contribute to the perception of an animate agent creating a sound. Nielsen, et al. conducted an auditory analogue of [5], using synthesized mosquito sounds [11]. Using binaural spatialization software, the authors varied the direction of motion and velocity of the mosquito sound and asked participants to rate how likely the sound was being produced by an animate source. Interestingly, velocity changes were significantly rated as more animate when compared to the other paradigms of manipulation including directional change of motion or no change at all. It gives rise to the question of what aspects of sound, and its organization in time, can be manipulated to change the perception of animacy in music.

Perhaps the closest analogue to changes of direction, motion, and intentionality in music would be the use of expressive timing. In music, expressivity is regarded by musicians to be the most important aspect of performance characteristics [12], and in performance, emotions are expressed through subtle implicit deviations in timing, dynamics, and intonation. In one study, Gabrielsson and Juslin asked participants to perform a melody to express different emotions such as sadness, happiness, and anger, as well as to perform with "no expression" [13]. The fluctuations in performance timing strongly differed from performances of other emotions. Moreover, timing variations were the most salient cues for an expressive musical performance, and participants that were asked to perform with "no expression" contained the smallest deviations in timing. Juslin defines expressivity as "random variations that reflect human limitations with regard to internal time-keeper variance and motor delays," [14]. Geringer et al. expands upon this, discussing the role of consistent expressivity as necessary for the appreciation of music as human-produced [15].

Manipulations in timing appear to elicit the greatest amount of change in the perception of human character in music. Johnson et al. asked participants to rate their perception of how "musical" a piece of music was (specifically a performance of Bach's Third Cello Suite), and found that when the music was manipulated to contain an exaggerated amount of rubato, "musical" ratings trended downwards [16]. However, as rubato was manipulated to contain less than the original performance, participants perceived the music as significantly less "musical". Bruno Repp analyzed the relationship between microtiming variations and larger scale timing variations found in rubato [17]. When participants were asked to perform Chopin's Preludes No.15 and No.6 with a "normal" expression metronomically (without aid of metronome), and in synchrony with a metronome, Repp discovered that articulated rubato performances of normal expression were simply exaggerated gestures of the random variations of timing that naturally occur due to human limitations. This suggests that the expressive nature of rubato mimics the micro-timing variations that already exist. This provides a stable foundation for the basis of the present study.

This study aims to elucidate the perceptual mechanisms involved in the perception of an animate human performance or an inanimate computer MIDI sequence using micro-timing variations. If microtiming variations are a salient cue for the perception of animacy in music, where is the tipping point between the perception of a "deadpan" computer generated MIDI sequence or an emotionally stimulating human performance? Furthermore, what is the optimal magnitude of variation that induces a convincingly human performance?

Using two experimental paradigms, this study explored how the perception of animacy is affected as a computer MIDI sequence is applied with variance, and from the other direction, how the perception of animacy in a raw human performance changes as microtiming variations are decreased. We hypothesized that animacy in music is most likely not linear, meaning that participants do not hear "aliveness" on a sliding scale, and that there is a likely "sweet spot" in rubato where performances that are too straight will be thought of as robotic, as will pieces that have too much variance.

II. Experiment 1: Applying Variance to Computerized Performances

Our first study used sequenced performances, adding varying levels of variance to each performance. As this study was converting "robotic" performances to more "human" performances, we advertised it as "Rohum" for participants (as a way of keeping it separate from "Humbot", which altered human performances).

A. Methods

Participants. Students of the Louisiana State University (LSU) School of Music (N=38) were recruited for participation (18 Females, 20 Males, mean age: 20.4). Experimental trials were conducted in the Music Cognition and Computation Lab.

Stimuli. Bach's Toccata and Fugue in D minor (BWV565) and Bach's Concerto for Oboe and Violin (BWV1060a) were computer generated using Finale MIDI music notation software (MakeMusic) to produce precisely metrically organized note onsets. Excerpts were split evenly by time (5 seconds each). Using Logic Pro X (Apple Inc.), tempo, timbre, and velocity fluctuations were minimized in order to eliminate any effect of expressiveness outside that of timing. Using Max/MSP software (Version 7; Cycling '74), we applied variance to each recording by randomly adding or subtracting an interval of time determined by dividing a maximum note displacement time of 500ms into 100 equal windows. For example, onsets occurring at 500ms and 1000ms with 5ms variance would be manipulated by randomly adding or subtracting 5ms to create a new manipulated onset of either 495/505ms or 995/1005ms respectively. If the recording were set to include a variance of 10ms, the same excerpt's onsets would be altered so that they fell at either 490/510ms and 990/1010ms after the beginning of the recording.

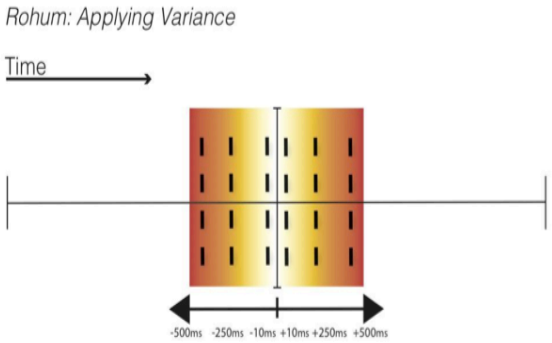

Fig. 1 Experiment 1: Rohum. Variance was applied to each recording by randomly adding or subtracting an interval of time on the order of milliseconds to all note onsets. The amount of variation was therefore fixed and determined by dividing a maximum offset time of 500ms into 100 different degrees of variance increasing at 5ms increments. Centered line represents metric beat; vertical dashed lines represent manipulated note onset times, shifting away from the beat, as variance increases.

Design. Participants were first asked to complete the Goldsmith Musical Sophistication Index, in order to later examine the relationship between ratings and relative levels of musicality [18]. After completing this survey, participants were split into two conditions, each with three blocks of ratings, and each played approximately 50 recordings of MIDI performances (~ 10s long). Participants were told that some of the performances were played by humans and others were sequenced, and were asked "how alive does this performance sound?". Each recording was rated on a 7-point Likert scale (from 1, definitely not alive, to 7, definitely alive).

B. Results

Animacy ratings. The animacy ratings were associated with the degree of variance of each stimulus. As mentioned above, we hypothesized that too much variance would lead to decreased perceptions of aliveness, but that too little variance would evoke a similar response. There is likely an ideal level of fluctuation that would create a sense of animacy. The hypothesis of a significant arc, however, was not supported. As can be seen with in Figures 2 and 3, however, the data fits nicely with a linear model. When fit with a linear regression, there was a significant (negative) correlation between the amount of variance applied and the perception of "aliveness". Results from both BWV565 and BWV1060A yielded significant results ($p<.001$), but with a relatively small effect size.

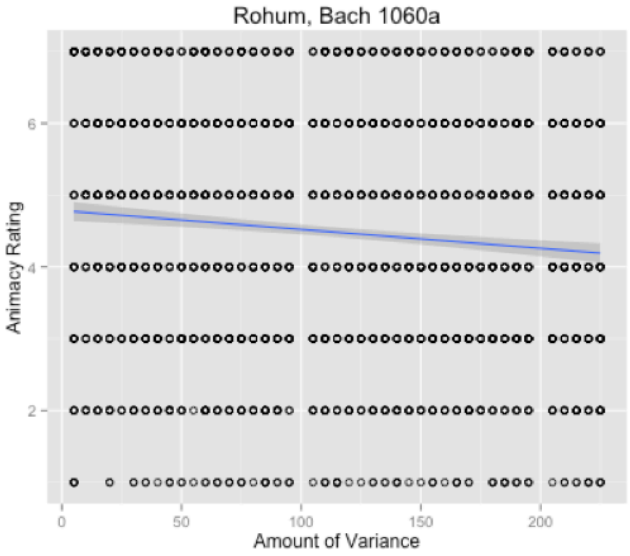

Fig. 2 Linear regression model demonstrating the relationship between the animacy rating of BWV1060a (on the y-axis) and the amount of variance ($p < .001$; small effect size)

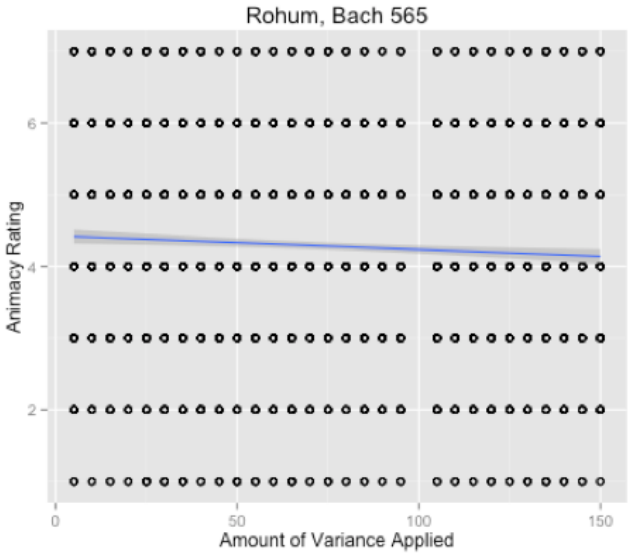

Fig. 3 Linear regression model demonstrating the relationship between the animacy rating of BWV565 (on the y-axis) and the amount of variance ($p < .001$; small effect size)

C. Discussion

The results indicate that even a slight increase in amount of variance makes the music sound more unnatural. The animacy ratings seemed to taper down as variation in inter-onset intervals increased. Increased microtiming variations using these methods did not seem to emulate a real human performance. The results here might also be seen as supporting the findings of Johnson et al. which, as discussed above, found that exaggerated rubato negatively affected the perceived musicality of a performance [16]. In the current study, as rubato increased, "aliveness" went down, which could be seen as a proxy for musicality, in some ways.

III. Experiment 2: Quantizing a Human Performance

The purpose of the second experiment ("Humbot") was to approach the question of how animacy ratings change with manipulation of microtiming variations, but from the opposite direction of the previous experiment ("Rohum"). This study is comprised of both a lab and a web study. The latter study was originally meant as a way of replicating the former study. Ideally, we would see an arc of "animacy ratings", based on the human performances that were quantized to varying degrees.

A. Methods

Stimuli. Humbot. A pianist performed Bach's Concerto for Organ in G major (BWV592) and Chopin's Mazurka 49 in F Minor (Op.68, no.4) on a MIDI keyboard. The pianist was instructed to perform the pieces naturally. MIDI recordings of both performances were divided into 10 degrees of quantization strength. Using Logic Pro X software (Apple inc.), tempo, timbre, and velocity fluctuations were minimized, and quantization strength was systematically applied on a continuum from 0% to 100% quantization by increments of 10%. Both recordings were quantized to the nearest 16th note. Thus, natural variation found in human performance decreased as quantization strength was increased.

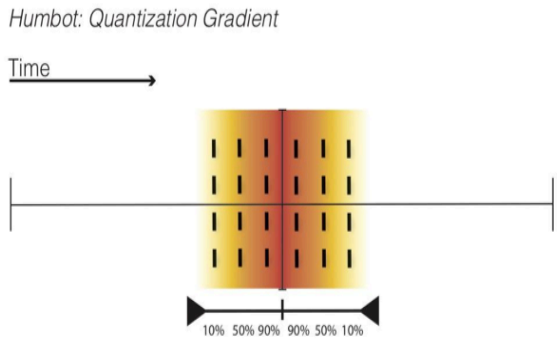

Fig. 4 Experiment 2: Humbot. Natural variance due to human limitations in performance were systematically reduced along a quantization gradient. Recordings were divided into ten different degrees of quantization strength where raw human performance was manipulated to increase in quantization strength from 0% quantization to 100% quantization by increments of 10%. Centered line represents metric beat; vertical dashed lines represent natural human fluctuations of timing in performance, shifting closer to the metric beat, as quantization strength increases.

B. Experiment 2a: Lab Study

Participants. Eighteen participants of the LSU School of Music (10 Males, 8 Females, mean age: 20.6 years, SD= 1.3) were recruited, and were given course credit for their participation. There was an overlap between participants recruited for Experiment 1 and Experiment 2 due to convenience sampling. Experimental trials were conducted in the Music Cognition and Computation Lab at LSU.

Design. Nearly identical to the first experiment, participants were asked to complete the Goldsmith Musical Sophistication Index, and were instructed to listen to approximately 50 recordings (~10-20s long) at degrees of quantization that were randomly selected. Participants rated each recording on a 7-point Likert scale (from 1, definitely not alive, to 7, definitely alive).

C. Experiment 2b: Web Study

Participants. Twenty-nine volunteer participants (14 Males, 15 Females) were recreuited through various social media platforms, such as Facebook, Twitter, and Reddit. Participants were not required to be a part of a university setting, and we did not require previous musical training.

Design. All participants were allowed to access the study from any location. Participants were given a link and voluntarily completed the study using various web browsers. Participants were prompted on the screen to listen to 5-10s recording excerpts at randomly selected degrees of quantization strength. After each stimulus, participants were asked to use the number keys 1 - 7 to rate each recording (from 1, definitely not alive, to 7, definitely alive) before moving on to the next recording. The participant had the choice of how quickly to move through the experiment.

D. Results

Animacy ratings. In order to examine the analogue between visual and auditory animacy, the results were analyzed in a method similar to Looser and Wheatley [8]. Like their study, we analyzed animacy ratings by first transforming them linearly to a scale from 0 to 1 (0 = less likely to performed by a human, 1 = more likely human). In order to find the point at which a recording was equally likely to be perceived as animate or inanimate—this point is known as the point of subjective equality (PSE; [8])— the animacy data from each participant was fit with a cumulative normal function across all recordings. A t-test was then performed between the PSE of each recording and the midpoint of the transform. The results of this t-test were significant ($p < .001$), indicating that the transformation from one end of the "aliveness" spectrum is not linear.

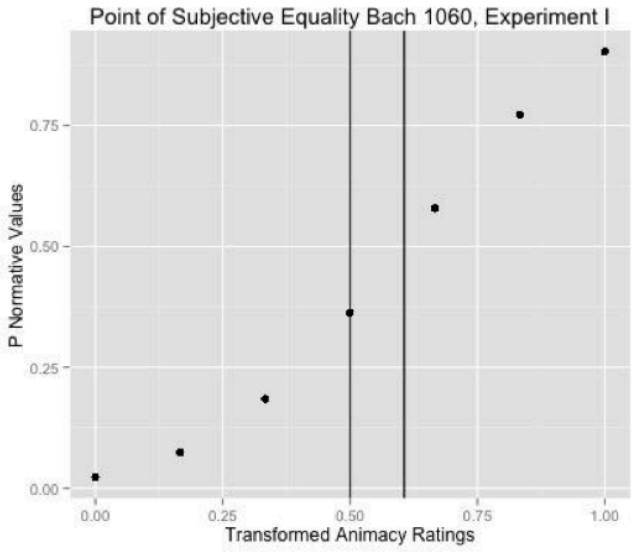

Fig. 5 Experiment 2a, the point of subjective equality for the animacy ratings of Bach's BWV1060a. This indicates that the relationship between the two extremes on the alive/not-alive spectrum do not follow a linear relationship.

E. Discussion

Ideally, we would be able to demonstrate that an arc of "aliveness" exists depending on the level of quantization. We were unable to find such a result with the lab study, which we think might be largely due to the fact that we had a relatively small sample size. As such, we decided to use the web study (discussed below) less as a replication, and more as a way of increasing statistical power.

The web study results did, in fact, indicate a trend toward an optimal level of quantization strength, where highest animacy ratings were weakly correlated. As can be seen in Figure 6 (below), we find that the original performances were correctly rated as "alive", but that these ratings actually increased as a minimal level of quantization was applied. Interestingly, however, these ratings sharply declined as quantization went past a threshold (around 50%). This effect seems to exist most strongly in the Chopin pieces, and less so in Bach, indicating a top-down style-processing component is at play.

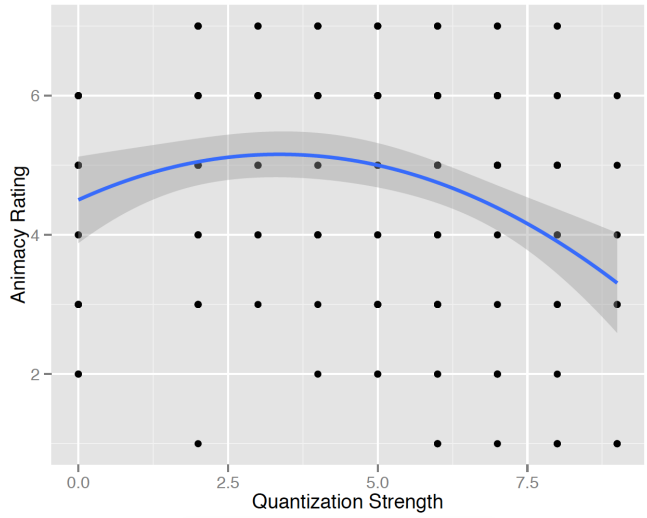

Fig. 6 Chopin's Mazurka 49 in F Minor (Op.68, no.4) from the web-study, indicating that there is likely a point at which there is an optimal level of quantization in rubato. As a corollary to Experiment 1, where too much rubato hindered the perceived animacy, it seems that too much quantization has a similar effect.

IV. Conclusion

In Experiment 1 (Rohum), we hypothesized that increasing variance would result in higher animacy ratings. However, the contrary was found. As variance was introduced, animacy ratings promptly decreased, suggesting that the method of applying variance may have been too rigid to convincingly emulate human performance. Variance was applied by randomly selecting between adding or subtracting an interval of time that was a multiple of 5 in the range of 0 to 500ms. That is, no random jitter was applied within each interval of time, simply, the original note in the music was shifted by a fixed amount of time to randomly fall either before or after the metric beat. Evidence from Experiment 1 suggests that this model for synthesizing natural variation due to human limitations does not communicate a convincing perception of human performance. The unpredictability of random variations makes every song unique, and does contribute to the 'living' character in music [14], but the present study suggests that the perceptual mechanisms involved in discerning between animate/inanimate music are highly sensitive to the type of microtiming variations that are embedded within the music.

Future work will also examine the types of variance and fluctuations that might be applied. Noise and random fluctuations come in a variety of forms, ranging from "white" noise (uncorrelated variance) to "brown" noise (1/f2 noise) [19]. Human performance has been shown to be a mixture of "pink" (1/f noise) and "white" noise [20], [19], [14], [21], [22]. Slower tempos, or larger intervals between beats produces a variance that more corresponds to 1/f noise. A task testing pure reaction time, such as asking participants to press the spacebar as quickly as possible will produce a variance of reaction time intervals that is more associated with white noise [21]. Preference for 1/f noise variance has been found in computer-sequenced music [23], however, the question of how much variance is necessary to produce a convincingly human performance, and if it is a categorical perception, still remains. Experiment 1 of the present study aimed to produce convincing human performances given computer-generated music as a starting point and applying "fixed" variance. Future research would test if pure white and pure 1/f noise produce significantly higher animacy ratings with increased application of variance, compared to the "fixed" variance method. Furthermore, if a PSE is found, employing a same/different task over the PSE would shed light on if the perception of animacy in music is categorical, thus suggesting consistency with the categorical perception of animacy in faces [8].

The results indicate that the manipulation of rubato using variations in microtiming has some effect on perceived animacy. The size of the effect, however, seems to be composer and genre specific. This is consistent with previous literature that suggests we are more likely to perceive objects as being animate than inanimate as evolutionarily, this is best for survival. Interestingly, this perceptual mechanism also extends into the auditory domain as it relates specifically to music perception [8]. The web study results revealed a trend towards an optimal level of quantization strength in the Chopin stimuli where animacy ratings initially increased in an inverse-U shape, reaching an apex point, but then promptly began to drop down as quantization strength continued to increase. We would argue that these minor changes in microtiming serve as a way to display intentionality and are communicative of human performance. Together, these studies illustrate how the perception of animacy in music can be manipulated by systematically adjusting microtiming variations, and perhaps more importantly, how the method of manipulation (Rohum vs. Humbot) affects the perception of animacy. Future work will continue along this train of thought, engaging in more analyses of stylistic differences and the perception of animacy, as well as the specific points at which rubato is applied, and how the shifting of such points affects the perception of aliveness.

REFERENCES

[1] F. Heider, M. Simmel, "An Experimental Study of Apparent Behavior," *The American Journal of Psychology*, vol. 57, pp. 243–259, 1944.

[2] J. A. Stewart, "Perception of animacy," Unpublished Psychology, University of Pennsylvania, 1982.

[3] D. Premack, "Intentionality - how to tell mae west from a crocodile - a review of intentional stance," *Behavioral and Brain Sciences*, vol. 11, 1988.

[4] D. Premack, "Cause/induced motion: Intention/spontaneous motion," *Origins of the Human Brain*, vol. 321, 1995.

[5] P.D. Tremoulet, J. Feldman, "Perception of animacy from the motion of a single object," *Perception,* vol.29, 2000.

[6] B.J. Scholl, P.D. Tremoulet, "Perceptual causality and animacy," *Trends in Cognitive Sciences*, vol. 4, issue 8, 2000.

[7] G.J.Y. Broze III, "Animacy, anthropomimesis, and musical line," Unpublished Graduate Program in Music Theory, The Ohio State University, 2013.

[8] C.E. Looser, T. Wheatley, "The Tipping Point of Animacy: How, When, and Where We Perceive Life in a Face," *Psychological Science*, vol. 21, issue 12, pp. 1854–1862, 2010.

[9] T. Wheatley, S.C. Milleville, A. Martin, "Understanding Animate Agents: Distinct Roles for the Social Network and Mirror System," *Psychological Science*, vol. *18* issue 6, pp. 469–474, 2007.

[10] E. Halgren, P. Baudena, J.M. Clarke, G. Heit, K. Marinkovic, B. Devaux, et al., "Intracerebral potentials to rare target and distractor auditory and visual stimuli. II. medial, lateral and posterior temporal lobe," *Electroencephalography and Clinical Neurophysiology*, vol. 94, issue 4, 1995.

[11] R.H. Nielsen, P. Vuust, M. Wallentin, "Perception of animacy from the motion of a single sound object," *Perception*, vol. 44, issue 2, 2015.

[12] E. Lindström, P.N. Juslin, R. Bresin, A. Williamon, "'Expressivity comes from within your soul': A questionnaire study of music students' perspectives on expressivity." *Research Studies in Music Education*, vol. 20, issue 1, 2003.

[13] A. Gabrielsson, P.N. Juslin, "Emotional expression in music performance: Between the performer's intention and the listener's experience," *Psychology of Music*, vol. 24, issue 1, pp. 68-91, 1996.

[14] P. Juslin, "Five Facets of Musical Expression: A Psychologist's Perspective on Music Performance," *Psychology of Music*, vol. 31, issue 3, pp. 273-302. 2003.

[15] J. M. Geringer, J.K. Sasanfar, "Listener perception of expressivity in collaborative performances containing expressive and unexpressive playing by the pianist," *Journal of Research in Music Education*, vol. 61, issue 2, 2013.

[16] C.M. Johnson, "Effect of rubato magnitude on the perception of musicianship in musical performance," *Journal of Research in Music Education*, vol. 51, issue 2, 2003.

[17] B. Repp, "Relationships Between Performance Timing, Perception of Timing Perturbations, and Perceptual-Motor Synchronisation in Two Chopin Preludes," *Australian Journal of Psychology,* vol. 51, issue 3, pp. 188-203, 1999.

[18] D. Müllensiefen, B. Gingras, J. Musil, L. Stewart, "The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population," *Plos One*, vol. 9, issue 2, 2014.

[19] P.N. Juslin, A. Friberg, R. Bresin, "Toward a computational model of expression in music performance: The GERM model," *Musicae Scientiae*, vol. 5, 2002.

[20] R. F. Voss, J. Clarke, "1/fnoise in music and speech," *Nature*, vol. 253, 1975.

[21] D.L. Gilden, T. Thornton, M.W. Mallon, "1/f noise in human cognition," *Science*, vol. 267, 1995.

[22] D.L. Gilden, "Cognitive emissions of 1/f noise," *Psychological Review*, vol. 108, issue 1, 2001.

[23] H. Hennig, R. Fleischmann, A. Fredebohm, Y. Hagmayer, J. Nagler, A. Witt, et al., "The nature and perception of fluctuations in human musical rhythms," *Plos One*, vol. 6, issue 10, 2011.