Effects Related to Synchrony and Repertoire in Perceptions of Robot Dance

Eleanor Avrunin
Dept. of Computer Science
Yale University
eleanor.avrunin@yale.edu

Justin Hart
Dept. of Computer Science
Yale University
justin.hart@yale.edu

Ashley Douglas
Dept. of Theater Studies
Yale University
ashley.douglas@aya.yale.edu

Brian Scassellati
Dept. of Computer Science
Yale University
scaz@cs.yale.edu

ABSTRACT

In this work we identify low-level aspects of robot motion that can be exploited to create impressions of agency and lifelikeness. In two experiments, participants view split-screen videos of multiple robots set to music and rate the robots on their dance ability, lifelikeness, and entertainment value. The first experiment tests the impact of the correspondence (or lack thereof) of the robot's motion to the underlying rhythm of the music, and the effect of matching changes in the robot's movement to changes in the music, such as a phrase of vocals or drumming. This motivates a second experiment which more deeply explores the relationships of asynchrony and changes in motion repertoire to participants' perceptions of the lifelikeness of the robot's motion. Findings indicate that perceptions of the lifelikeness of the robot and the quality of the dance can be manipulated by simple changes, such as variation in the repertoire of motions, coordination of changes in behavior with events in the music, and the addition of flaws to the robot's synchrony with the music.

Categories and Subject Descriptors

I.2.9 [**Artificial Intelligence**]: Robotics

General Terms

Experimentation, Human Factors

1. INTRODUCTION

Regular, rhythmic motion is common to mechanical devices, from metronomes to assembly-line robots, yet, in the form of dance, it is also a critical and expressive part of human culture and interaction. To a casual human observer, the difference between the two seems obvious, but providing a comprehensive definition is difficult even for experts on the topic [3]. One factor in this difference may be that observers assume that a human dancer is an agent with motivations that drive his or her motions, while a factory robot is seen purely as a machine carrying out a programmed series of steps. While a dancer's intent cannot be directly observed, it can be inferred from his or her movements, and a human audience is accustomed to making such attributions. In this work we investigate how low-level aspects of a robot's motion can make use of that habit of inference to affect perceptions of lifelikeness and agency.

In prior work on agency in our group [11], deviations from the expected course of action during a game of rock-paper-scissors led to attributions of agency to a humanoid robot. Here, we shift to the context of dance and explore control mechanisms to produce such attributions at a lower level. Where the previous study used a cheating action to produce cognitive attributions of intent, we exploit observers' perceptual expectations of dance to achieve an increased impression of lifelikeness and agency with small changes in the robot's motion. We find that simple mechanisms such as manipulating the robot's synchrony with the rhythm of the underlying music, changing its repertoire of moves, and synchronizing the robot's motion with events in the music can affect perceptions of lifelikeness, dance quality, and entertainment value.

Dance and rhythm have recently been studied extensively in the robotics literature and used as a medium for human-robot interaction. Michalowski, Sabanovic, and Kozima used a free-form dancing task to study the role of rhythm in social interactions [9]. Work has been done in building robots to dance independently [6] or as a partner to a human dancer [12]. There has also been considerable work in music performance and rhythmic synchrony [2, 14].

Here we are concerned with what aspects of a robot's dance can make it appear lifelike. If observers have expectations of what a programmed machine will do in a certain situation, deviating from those expectations may make a robot appear more like an agent and less like a machine. Camurri, Krumhansl, Mazzarino, and Volpe [1] demonstrated that observers watching a dance form expectations regarding the dance, anticipating a performer's movement. A mild form of such an expectation is for the observer to follow the beat of the music and expect the dancer's motion to be synchronized with it. A robot's deviation from the beat, or

HRI'11, March 6–9, 2011, Lausanne, Switzerland.

from other such expectations, may appear to be a "mistake", such as might be made by a human dancer.

In contrast, observers also form different expectations of dancers as agents, and conforming to the patterns of motion expected of an agent may make a robot appear more lifelike. Krumhansl and Schenck [7] show that the boundaries of sections in a dance performance and the music accompanying it are tightly correlated. Observers watching the dance without the music, or listening to the music without the dance, mark boundaries in similar locations. Human dancers listening to the music may anticipate or react to the musical change with a shift in their performance. A robot behaving as would be expected of a dancer in these situations may help create the perception of agency, just as would executing social cues in a direct interaction.

In this work, we attempt to use these expectations to produce attributions of agency. Since participants' responses to questions directly asking if the robot is an agent or is "trying to dance" would confound their attitudes toward robots and their reactions to our stimuli, we instead ask them to rate the robot's performance on aspects such as "Dance quality" and "Lifelikeness". We expect ratings of "Dance quality" to reflect factors such as the robot's ability to maintain the beat of the music, while "Lifelikeness" will reflect similarity to human dancers. Other treatments of lifelike motion have focused on aspects such as actuation, control, or the measurement of the motion of living things [5,8,13]. In contrast, we are interested in using perceptions of lifelikeness as a reflection of attributions of agency by our study participants. Our goal is to investigate characteristics of the robot's motion that may lead participants to make these attributions.

2. EXPERIMENT I

The first experiment attempts to identify high-level qualitative characteristics relevant to our concept of lifelike motion. To this end, the following hypotheses are tested:

Hypothesis 1: Motion that is "on" the musical beat will be perceived as better dancing.

This follows from the notion that study participants will judge the quality of the robot's dances much as they would the quality of human dances, and will consider synchrony with the music to be an important criterion.

Hypothesis 2: Robots which move out of sync with the music will be perceived as worse dancers and less lifelike than robots which move in sync with the music.

If perceptions of high-quality dance and lifelike motion are simply a function of synchrony between the timing of the onset of motion and the beat of the music, then they will not be created when that synchrony is absent.

Hypothesis 3: Participants will find both better dancing and more lifelike motion to be more entertaining.

Just as audiences watch professional dancers for the sake of their talent, they find lifelike animatronics, such as those found in theme parks, to be amusing.

Hypothesis 4: Changes in the robot's selection of dance moves that are correlated to events in the music will be rated more positively than ones that do not correlate to changes in the music.

Performers often synchronize dance steps to events, such as cymbal crashes or changes in musical mood, and composers scoring music for movies will synchronize musical sounds to events in the film. This hypothesis predicts the impact of this synchrony.

Three additional characteristics of dance were identified for testing in the first experiment. Due to space considerations, this section of the paper focuses on the trials which provide the clearest and most interesting results, omitting three trials dedicated to testing these additional hypotheses.

2.1 Method

2.1.1 Participants

The experiment was structured as an online survey. We recruited 200 participants (133 female, 67 male), between the ages of 18 and 65, through social networking sites. Participants were not compensated for participation in the study. Only participants who completed the full survey, including demographic information, are included in the results.

2.1.2 Robot

The robot used in this experiment is Keepon, Figure 1. Keepon is a small creature-like robot with a snowman shape, designed to study child social interaction [9]. It sits atop a black cylinder containing four motors, a PID controller, and a motor driver. Its body has a soft rubber exterior, with two cameras in the eyes and a microphone in the nose. Keepon's body has four degrees of freedom: tilting forward and backward, rotating left or right, leaning side to side, and bouncing up and down. The PID controller can set parameters for acceleration of the motors to the desired position [9].

Keepon's body, like that of many animated characters, abstracts away many of the complex degrees of freedom and appearance characteristics of humans or real-life animals that many other robots attempt to model. This simplicity allows the present study to focus on overall characteristics of motion, rather than specifics such as gaze control, head pose, or the placement of limbs, which would become relevant on more complex platforms.

2.1.3 Procedure

On the website, participants are first presented with an informed consent agreement that must be accepted in order to participate, then a page of instructions explaining the experiment. Participants are informed that the purpose of the study is to explore robotic dance and lifelike motion, that participation is voluntary, and that they will not be compensated for participation.

Participants are then presented with videos of three or four robots moving in pre-recorded motions, shown in split-screen format, as seen in Figure 1. The split-screen presentation allows participants to easily make comparisons across conditions. Participants are able to replay the videos, accounting for the possible issue of not being able to concentrate on all conditions simultaneously. The videos are between 30 and 90 seconds long, set to clips of popular drum- or bass-heavy music. The order that the robots appear from left-to-right in the videos is randomized to avoid any possible ordering effects. To disambiguate the videos from each other, the labels "Keepon A," "Keepon B," and "Keepon C" are placed under the videos on the website, which records which video is 'A', 'B', or 'C' for each participant. The same robot is used in all videos. While the robot's motions are pre-programmed as Processing [10] and Java programs, the videos were manually reviewed and synchronized to the music in order to visually confirm consistency. Each trial (listed in Section 2.1.4) is presented on its own page. Participants

Figure 1: Example screen capture of video presented to participants in the experiment. Each split-screen video is labeled "Keepon A," "Keepon B," or "Keepon C" in left-to-right order under the video on the website.

are not allowed to go back and re-answer questions from a trial after proceeding to later trials.

Under each set of videos are a set of Likert-scale type questions and a free response question. Upon completing the trials, participants are presented with a page of overall questions, a page of demographic information, and a page thanking them for participating and providing contact information should they have any questions or requests.

2.1.4 Trials

Schematics of the dances in these trials are in Figure 2.

Trial 1 - Synchronization of movement with musical rhythm: The first trial assesses the relevance of correspondence between the robot's dance rhythm and that of the music to our measured variables. We define a move to be leaning, bouncing, leaning and bouncing simultaneously, or similarly coordinated motions. For this test, moves consist strictly of side-to-side tilting motions across the robot's sagittal axis. The robots in the split-screen video represent three experimental conditions: *On the beat*, in which the robot's moves match the beat of the music; *Half-on/half-off*, in which the robot's move is on the beat half of the time, and off the beat half of the time; and *Off the beat*, in which the tempo of the robot's motion does not quite match that of the music, and the moves are consequently not synchronized with the beat of the music.

Trial 2 - Change in music and motion: The second trial evaluates the relevance of synchronizing change in motion with noticeable changes in music. The robot alternates between three moves: bouncing, scrunching and tilting to the side, and a combined lean and bounce. The music alternates between vocal and percussion sections every four measures, each measure being two beats of music. The three conditions for this test are changing moves every three, four, or five measures, such that the condition in which the robot changes its dance move every four measures does so in sync with the musical change. In all conditions, the robot's motion is synchronized with the beat of the music.

2.1.5 Measures

Each trial contains three 7-point Likert-scale questions, on which participants rate each robot in the split-screen video:

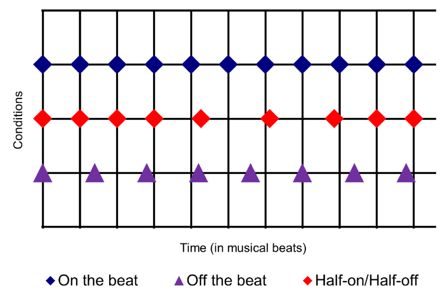
Please rate how well each Keepon dances, where a score of 1 is labeled *Very poorly*, 4 is labeled *Neither well nor poorly*, and 7 is labeled *Very well.*

Please rate how lifelike each Keepon is, where a score of 1 is labeled *Very mechanical*, 4 is labeled *Neither lifelike nor mechanical*, and 7 is labeled *Very lifelike.*

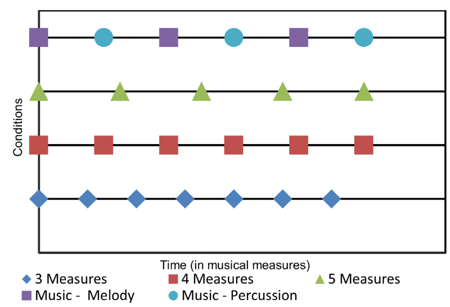
Please rate how entertaining each Keepon's dance is, where a score of 1 is labeled *Very boring*, 4 is labeled *Neither entertaining nor boring*, and 7 is labeled *Very entertaining.*

Additionally, each trial has the free-response question, *How do these robots differ from each other?*

A response is required for each question.



(a) Trial 1.



(b) Trial 2.

Figure 2: Schematics of dances in Experiment I. In Trial 1, *On the beat* is consistently with the music, *Off the beat*'s beat is slower and does not match the music, and *Half-on/half-off* is synchronized half of the time, with sporadic offsets taking it off-beat. In Trial 2, the top line illustrates the sound of the music. The other lines illustrate points at which the robot changes the move it performs in each dance.

2.2 Results

Hypothesis tests are performed using repeated measures Analysis of Variance (ANOVA). Tests of within-subjects effects use standard corrections to the degrees of freedom applied in cases where Mauchly's test statistic, W, is significant ($p < 0.05$), indicating that the sphericity assumption is violated. These adjustments make the estimate of the F-score more conservative. For $\epsilon > 0.75$, we used the Huynh-Feldt

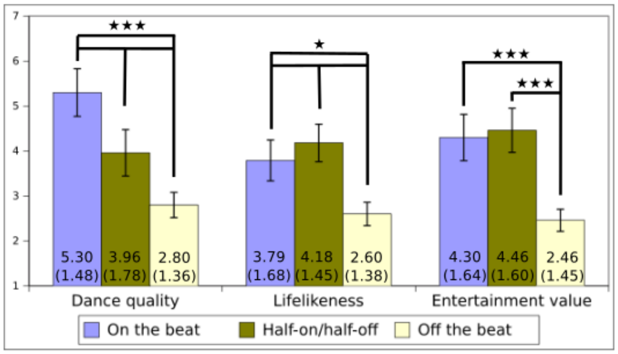


Figure 3: Trial 1, effects of synchronization of movement with musical rhythm. While the robot that synchronizes its movement with the music (*On the beat*) is rated as the best dancer, the robot that is only partially synchronized (*Half-on/half-off*) is rated as more lifelike. Error bars on all charts in this paper indicate standard error. The following symbols are used to indicate statistical significance $\star : p \leq 0.05$, $\star\star : p \leq 0.01$, $\star\star\star : p \leq 0.001$. Numbers at the bottom of the bars indicate the means, numbers in parentheses indicate standard deviation.

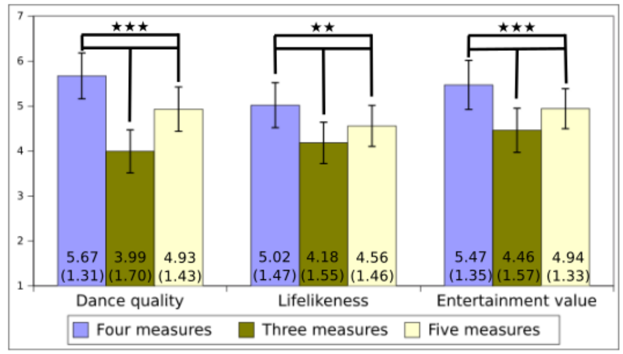


Figure 4: Trial 2, change in music and motion. The robot whose moves change with changes in the music (*Four measures*) is rated as a better dancer, more lifelike, and more entertaining than the others.

correction, and for $\epsilon < 0.75$, we used the Greenhouse-Geisser correction, following the guidelines provided by Field [4]. Correlations are computed as Pearson correlations.

2.2.1 Trial 1: Synchronization of movement with musical rhythm

In the first trial, the main effects for all three measures, "Dance quality" ($F(1.811, 199) = 133.990, p < 0.001$), "Lifelikeness" ($F(2, 199) = 81.970, p < 0.001$), and "Entertainment value" ($F(2, 199) = 146.913, p < 0.001$) are significant.

As can be seen in Figure 3, the *On the beat* condition is perceived as a better dancer than either *Half-on/half-off* or *Off the beat*, supporting Hypothesis 1. Interestingly, the results for "Lifelikeness" do not support Hypothesis 2. *Half-on/half-off* is perceived as more lifelike than *On the beat* or *Off the beat*. Respondents rate *Half-on/half-off* to be about as entertaining as *On the beat*. Both, however, are significantly more entertaining than *Off the beat*.

It is interesting that the *Half-on/half-off* condition is rated as more lifelike and approximately as entertaining as *On the beat*. Responses to the free-response question help to elucidate the matter. Some participants compare the *On the beat* and *Off the beat* conditions to metronomes:

- *[On the beat] and [Off the beat] are very much like metronomes [...] [Half-on/half-off], at first perception seemed to be faulty in its movements, but the more I watched the video, the more life-like it became and more entertaining as it changed up the pace and style of its back and forth rocking.*

Other participants do not even notice that the *Half-on/half-off* robot is off the beat:

- *[Half-on/half-off] dances like a person would switching up the moves*

Participants appear to appreciate the variation in the inconsistent robot's motion, even though this variation only comes in the form of being arrhythmic. In contrast, they punish the *Off the beat* robot and compare it to machinery in free-response descriptions. Perfect synchrony with music is aspired to in dance, as participants acknowledge when they rate *On the beat* the best dancer, but it is rarely achieved by living dancers. In this test, we see that some error can appear more lifelike than perfection.

2.2.2 Trial 2: Change in Music and Motion

In Trial 2, the main effects for all 3 Likert scale measures are significant ("Dance quality": $F(1.857, 199) = 101.248$, $p < 0.001$, "Lifelikeness": $F(1.929, 199) = 34.528, p < 0.001$, "Entertainment value": $F(1.801, 199) = 50.888, p < 0.001$), as are all mean differences at $p < 0.001$.

It can be seen in Figure 4 that the robot that changes its moves with changes in the music, *Four measures*, is perceived as the best dancer, the most lifelike, and the most entertaining, supporting Hypothesis 4. This affirms the current practice of artists such as animators, dancers, and film makers. The synchrony of the robot's motion with events in the music improves participants' ratings of its dance. Some participants notice this synchrony, or the jarring changes of the other two conditions:

- *[Four measures]'s movements seem more in sync with the changes in the music than [Three measures] or [Five measures].*

Some even explicitly attribute the synchrony to planning or response to the music:

- *[Four measures] seems to have a little choreographed dance thing going on.*

Interestingly, the *Five measures* condition is rated higher than *Three measures*, and some participants only notice *Three measures* as being out of sync or jarring:

- *[Four measures] reacts more to the music, [Five measures] has a more steady switch pattern, [Three measures] is out of sync.*

One possible explanation for this is that, because the sections of vocals and drum in the piece are 4 measures long, *Three measures* never performs the same move for the entire musical phrase.

2.2.3 Hypothesis 3

To test Hypothesis 3, responses to the three Likert-scale questions on Trials 1 and 2 were correlated to each other. Supporting this hypothesis, all positive attributes are positively correlated: "Dance quality" - "Entertainment value" $r(1200) = 0.670, p < 0.001$, "Dance quality" - "Lifelikeness" $r(1200) = 0.692, p < 0.001$, "Lifelikeness" - "Entertainment value" $r(1200) = 0.657, p < 0.001$.

3. EXPERIMENT II

To elucidate the results of Experiment I, Experiment II focuses more closely on three factors: changes from being off-beat to on-beat, changes in repertoire, and the correlation of a change in performance with a change in the music.

Hypothesis 5: Motion that changes from being out of sync with the musical beat to being in sync will be perceived as better dancing and more lifelike than motion that simply maintains the beat of the music throughout the dance.

Most dancing is intended to be "on" the beat of the music, but many dancers have difficulty maintaining that synchrony. This hypothesis predicts that participants will perceive this change as a correction, much like a human dancer might make.

Hypothesis 6: Rapid and distinct changes in rhythm from asynchrony to synchrony will be perceived as more lifelike than gradual changes.

Dancers often make sudden, intentional corrections. While a machine might adjust a parameter to minimize the difference between its motion and the beat of a piece of music, dancers often count the beats of the music, then fix their motion to be in sync with the music.

Hypothesis 7: Combinations of multiple changes, either in the form of changes in the music corresponding to changes in the dance, or combined changes in the dance, such as going from off beat to on beat combined with changes in repertoire, will be rated better than changes in isolation.

Testing this hypothesis will determine whether more positive ratings can be achieved by combining the techniques which yielded improvements in Experiment I.

3.1 Methodology

3.1.1 Participants

Experiment II is conducted via an online survey in the same format as Experiment I, using the same recruitment methods. We recruited 118 participants (99 female, 14 male, 5 who reported other or chose not to respond to this item), between the ages of 18 and 72. As in Experiment I, participants were not compensated, and only participants who completed the entire survey are included in the results.

3.1.2 Robot

Experiment II utilizes the same robotic hardware as Experiment I. The software, however, is updated to allow for greater precision in the timing aspects of the robot's control. This allows "on-the-beat" and "off-the-beat" to be defined in terms of a delay, in milliseconds, from the beat of the music. Whereas Experiment I utilizes a hand-tuned program and careful observations of a classically trained dancer, Experiment II uses DJ mixing software. Moves are defined as taking a time period defined by the period of a quarter note, but limited by the speed of the robot's motion. In cases where the robot's motion is not fast enough to perform a move on the quarter note, the move is defined as taking a half note. This time period is tuned using the beats-per-minute of the musical track. Time offsets representing the quality of being "off-the-beat" are then algorithmically added.

3.1.3 Procedure

The format of the videos and survey used in Experiment II are the same as in Experiment I. The trials are always shown in the same order, as presented in Section 3.1.4, though the left-to-right split-screened order of the conditions is counterbalanced in each trial to avoid ordering effects, as before.

3.1.4 Trials

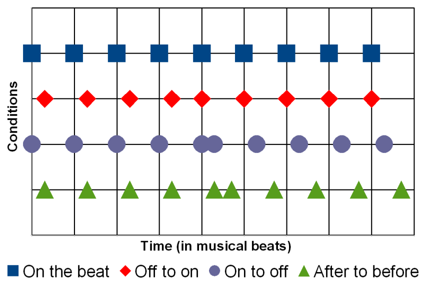
Schematics of the dances in these trials are in Figure 5.

Trial 3 - Change with respect to the beat of the music: Trial 3 tests Hypothesis 5, that a robot changing from dancing off the beat of the music to on the beat will improve participants' perceptions of its "Lifelikeness" and "Dance quality". The robot's moves are defined to take a half note (1120.34 milliseconds) to complete. For the purpose of this trial "Off the beat" dancing is defined as dance in which the move is performed 300 milliseconds after the beat. There are four conditions: *On the beat*, a control condition in which the robot's moves match the beat of the music; *Off to on*, in which the robot changes from dancing off the beat to on the beat; *On to off*, in which the robot changes from dancing on the beat to off the beat; and *After to before*, in which the robot goes from dancing 300 milliseconds after the beat to 300 milliseconds before the beat.

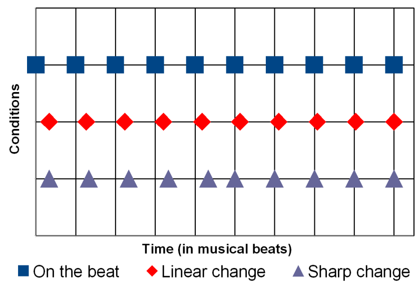
In all conditions of this trial, the robot performs a dance containing three moves: a lean and bounce to the left, the same motion to the right, and the same motion forward. In the three conditions in which the robot makes a change in its dance performance, the changes occur simultaneously, about 27 seconds into the song, on the 24th move.

Trial 4 - Sharp change versus smooth change in correspondence to the beat:. Trial 4 tests Hypothesis 6, that sharp changes in correspondence to the beat will be rated more positively than smooth changes. The robot's moves are defined to take a quarter note (649.28 milliseconds), and moves off the beat of the music are defined to occur 300 milliseconds after the beat. There are three conditions: *On the beat*, which is defined as in Trial 3; *Sharp change*, where the change from off-the-beat to on-the-beat proceeds as in prior trials, sharply, halfway through the clip; and *Linear change*, in which the offset for off the beat dance is linearly interpolated from 300 milliseconds at the start of the clip to 0 at the end.

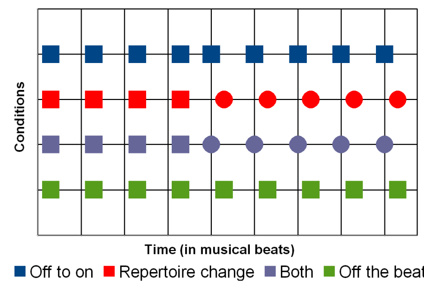
Trial 5 - Correlation to the beat versus repertoire: Trial 1 simply changes the correspondence of the robot's motion to the beat. Trial 2, in testing the correspondence of moves to musical events, also changes the robot's dance repertoire. This trial tests the relative effects of a change in correspondence to the beat of the music and a change in repertoire, and whether there is an interaction when both change simultaneously. The robot's moves are defined to take either a quarter note (492.00 milliseconds) or a half note (984.00 milliseconds), and the delay when the robot is the "off-the-beat" is 250 milliseconds. The first repertoire consists of two moves lasting a quarter note and one move lasting a half note, while the second repertoire has one quarter-note move and two half-note moves. In both cases, the dance consists of an eight-beat loop of four quarter notes



(a) Trial 3.



(b) Trial 4.



(c) Trial 5.

Figure 5: Schematics of dances in Experiment II. Vertical lines indicate the beat of the music. In Trial 3, the robots change their synchrony with the music halfway through the dance. In Trial 4, the *Linear change* robot interpolates its change, whereas the *Sharp change* robot changes halfway through. In Trial 5, either the robot's synchrony with the music, its move repertoire, or both change.

and two half notes. The song goes from the verse to the chorus 33 seconds into the clip, at which point the relevant change in the robot's behavior occurs. There are four conditions: *Off to on*, in which the robot makes a sharp change from off-the-beat to on-the-beat dance; *Change in repertoire*, in which the robot's selection of dance moves changes on the chorus, but the robot's motion is always off the beat; *Both*, in which both changes occur; and *Off the beat*, in which dance is always off the beat.

3.2 Results

3.2.1 Trial 3: Change with respect to the beat

Results for Trial 3 can be seen in Figure 6. The main effect is significant for "Dance quality" at $F(2.489, 137) = 8.023, p < 0.001$. It is also significant for "Entertainment value" at $F(2.771, 137) = 4.159, p = 0.008$. The main effect is nearly significant for "Lifelikeness" at $F(2.773, 137) = 2.041, p = 0.113$. Interestingly, the only important factor in participants' ratings of the robot's motion is whether or not it is always on the beat, and *On the beat* is the best performer on all measures. On dance quality, all mean differences with *On the beat* are significant at $p \leq 0.008$, but all others are not significant. Comparing this result to that of Trial 1, in which there is an obvious downward progression in dance quality as the robot is off the beat for more of the dance, it is unsurprising that the robot that is always on the beat is rated the best dancer.

What is more surprising is that there is no real difference between the performances of the remaining three conditions. No mean differences are significant for "Lifelikeness." While *On the beat* is predictably the best, Hypothesis 5 predicts that *Off to on* would be seen as an improvement or correction and rated positively by participants. Under this hypothesis, ending in sync with the music would be rated positively, and correcting incorrect dance would appear lifelike, giving *Off to on* a net improvement.

One possible interpretation for this result is that participants do not distinguish, or have no preference regarding the various changes made, and notice only that the robot is sometimes dancing off the beat. Under this interpretation any amount of time off the beat is enough to negate a period of being on the beat, so that all three conditions in which the robot is not perfectly on the beat are equally "bad." However, this interpretation conflicts with the results of Trial

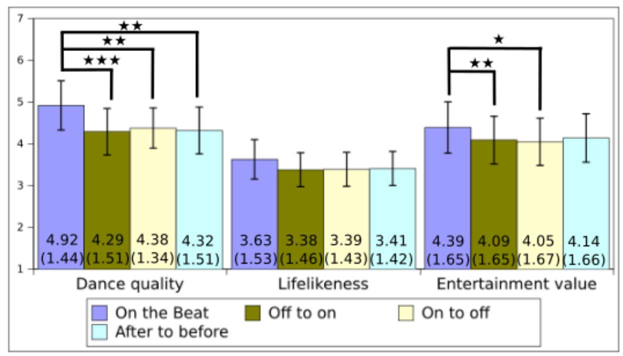


Figure 6: Trial 3, change with respect to the beat. The robot that is always synchronized with the musical beat (*On the beat*) is rated as a better dancer, more lifelike, and more entertaining than the others, which perform commensurately with each other.

1, in which *Half-on/half-off* is rated more lifelike and more entertaining than *On the beat*.

The crucial difference between being off the beat in Experiment I and Experiment II is how this behavior is defined and executed. In Experiment I, the robot is off the beat half of the time by delaying some moves and performing others too early. In Experiment II, being off the beat is programmed as a steady tempo and a strict offset from the correct beat. This offset is the same for every off-beat part of every condition, meaning that all conditions that are off the beat at any given moment are synchronized with each other. This either makes it easy for participants to rapidly pick out the condition that is on the beat, or creates confusion, as the off-the-beat robots are still synchronized. The uniform offsets from the beat may also be more difficult to perceive than the changing offset in Experiment I. The scores for all conditions in Experiment II, as can be seen in Figures 6, 7, and 8, are closer to 4, the middle of the 7-point Likert scale, than in Experiment I, which has greater variance. This suggests that even when participants recognize the difference they do not consider it a large one.

Future work will explore the use of random offsets from the beat, rather than constant ones. This differing offset and changing tempo will alleviate the problem of having several

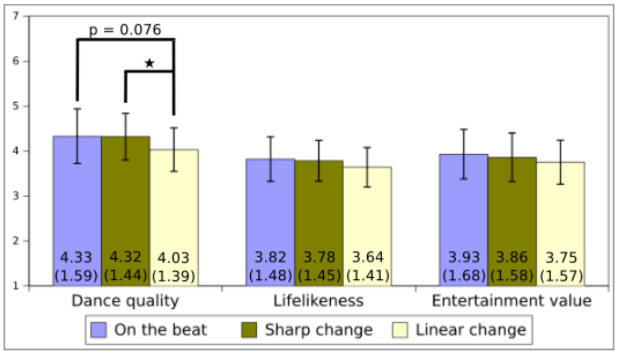


Figure 7: Trial 4, sharp versus smooth change. The robot that makes a rapid change from being off the beat to being on the beat (*Sharp change*) performs as well as the robot that is always synchronized with the beat (*On the beat*) on "Dance quality."

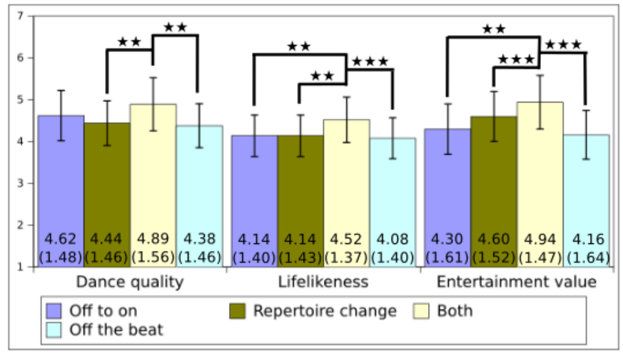


Figure 8: Trial 5, correlation to the beat versus repertoire. The robot that changes both repertoire and synchronization with the beat (*Both*) is rated as a better dancer, more lifelike, and more entertaining than the others.

robots which are off the beat of the music in sync with each other, and will more closely resemble the *Half-on/half-off* condition in Trial 1.

Given that *Off to on* does not gain an improvement in "Lifelikeness", it is unsurprising that the changes in behavior have a minimal effect on "Entertainment value," aside from *On the beat* performing the best. It does, however, support the hypothesis that the improvement in *Half-on/half-off* is due to enjoyment of the perception of lifelikeness.

3.2.2 Trial 4 - Sharp change versus smooth change in correspondence to the beat.

Results for Trial 4 can be seen in Figure 7. The main effect for "Dance quality" is significant at $F(1.805, 137) = 3.878, p = 0.026$, with "Lifelikeness" not significant at $F(1.772, 137) = 2.102, p = 0.132$, and "Entertainment value" nearly significant at $F(1.833, 137) = 2.974, p = 0.057$.

As can be seen in Figure 7, *Linear change* is rated the worst on all measures. Though the sums of the time delays between beat and move are the same for both *Sharp change* and *Linear change*, the *Linear change* condition makes a greater number of off-beat moves and spends a greater percentage of the song off the beat. One possible interpretation of the result is that participants weight these factors more heavily in their ratings than the total offset from the beat. Another possible interpretation is that it is simply harder to perceive the improvement in the linearly changing robot's motion. The change is very gradual, and participants may not recognize the slight differences between one beat and the next. A third possible explanation is that participants do perceive the change, but do not see it as a lifelike one. Human dancers often "catch the beat" of the music, while machines often gradually minimize the difference between their performance and optimal performance. It is possible that improving performance in this manner simply does not look lifelike to the participants.

Interestingly, the mean difference between *On the beat* and *Sharp change* for "Dance quality" is not significant. On this metric, participants rate *Sharp change* according to its performance at the end of the song. This is in contrast to Trial 3, in which *Off to on* performs commensurately with any robot which ever deviates from the beat of the music. This is further discussed in Section 3.2.3.

Recall that *Linear change* performs the worst, and is as bad as *Sharp change* at the start of the clip. Therefore, participants can detect the initial difference between *Sharp change* and *On the beat*, but appear to respond to the *Sharp change* condition's performance at the end of the clip. *Sharp change* does not, however, get the boost in lifelikeness that *Half-on/half-off* does in Trial 1, demonstrating that while inconsistency with the beat is lifelike, it is more lifelike if the asynchrony has some variance, as might a human dancer's errors, rather than being uniformly off the beat.

3.2.3 Trial 5 - Correlation to the beat versus repertoire.

In Trial 3, *Off to on* performs about as well as the conditions that make other changes, including *After to before*, which never dances on beat. In Trial 4, the performance of *Sharp change* is roughly commensurate with *On the beat*. A better understanding of this effect can be gained from Trial 5, where *Both* performs the best. Results for Trial 5 can be seen in Figure 8. All three main effects are significant ("Dance quality": $F(2.686, 137) = 4.958, p = 0.003$, "Lifelikeness": $F(2.808, 137) = 6.191, p = 0.001$, "Entertainment value": $F(2.016, 137) = 13.263, p < 0.001$).

In Trial 5, the condition that is rated the best is the one which not only changes from off the beat to on the beat, but also changes its repertoire of moves. Our original hypothesis for Trial 5 was that each change individually would improve performance slightly, and that, together, the benefit would be compounded.

On "Dance quality", the mean difference between *Both* and *Off to on* is not significant, indicating that participants identify that both of these robots finish the song on the beat. *Both* does, however, significantly outperform *Repertoire Change* and *Off the beat*.

Comparing the results of Trial 5 with the other two trials, it appears that the strategy of changing from off the beat to on the beat is most effective at improving lifelikeness if participants do not immediately perceive exactly what happened. In both Trial 3 and Trial 5, most of the conditions are synchronized for much of the song. This reduces the cognitive load in identifying the exact change that occurs when a condition switches to moving on the beat. Distinguishing between the conditions is a question of consciously noticing which robot appears different, and how. In both these cases, the single change from off the beat to on the beat does not produce a significantly greater perception of lifelikeness. These simple manipulations thus appear to work best if they are not easily spotted. Conscious recognition of the manipulation may be what eliminates its effect. In Trial 5, it may be the case that the condition that both catches the beat and changes its repertoire produces this effect because multiple changes occur at once. These changes are also accompanied by the change from the verse to the chorus of the song. Having so many changes happen concurrently may obfuscate the individual changes sufficiently for participants to perceive them unconsciously, rather than analytically, resulting in a perception of lifelike behavior.

4. CONCLUSION

In this work, we have programmed a creature-like robot, Keepon, to perform a number of dances in order to test hypotheses regarding the effects of different aspects of motion on observers' perceptions of the robot's lifelikeness and agency. We asked participants to rate the dance quality, lifelikeness, and entertainment value of different conditions. Hypothesis 1, that dance on the musical beat is perceived as superior dancing, is supported by this work. Hypothesis 3 is also supported, in that participants find better dancing and more lifelike motion to be more entertaining, as is Hypothesis 4, demonstrating the benefit of synchronizing dance events to musical ones.

Interestingly, Hypothesis 2 is not supported. The *Half-on/half-off* condition is the most lifelike and entertaining, suggesting that perfect synchrony is less lifelike than making mistakes. This leads to Hypotheses 5, 6, and 7. Hypotheses 6 and 7 are supported, as rapid, distinct changes and combinations of changes yield greater perceptions of lifelikeness. Hypothesis 5 is partly supported. Changes in synchrony are lifelike if either the performance is not uniformly off or on the beat, or if the change is masked, by being accompanied by other effects.

Creating the perception of a robot as a lifelike agent can be critical in human-robot interaction, but it is an extremely difficult task and how to go about it is not always clear. This work demonstrates that evaluations of the lifelikeness and quality of a robot's dances can be improved through a number of simple manipulations. We find that subtle mistakes and imperfections in the robot's motion, and responsiveness to changes in context, as in the case of the robot's dance changing with the music, can help the robot to appear more lifelike. The techniques outlined in this paper could be easily adapted to other robots, resulting in subtle, unexpected differences in their behavior which improve users' perceptions of them as lifelike entities.

Acknowledgments: This material is based upon work supported by the National Science Foundation under grants #0968538 and #0835767 and by the DARPA Computer Science Futures II program. The authors also acknowledge the generous support of Microsoft and the Sloan Foundation.

5. REFERENCES

[1] A. Camurri, C. L. Krumhansl, B. Mazzarino, and G. Volpe. An exploratory study of anticipating human movement in dance. In *Proc. Intl. Symp. on Meas., Analysis and Modeling of Human Functions (ISHF)*, pages 57–60, Genova, Italy, June 2004.

[2] C. Crick, M. Munz, T. Nad, and B. Scassellati. Robotic drumming: Synchronization in social tasks. In *Proc. 15th IEEE Intl. Symp. on Robot and Human Interactive Communication (RO-MAN 2006)*, pages 97–102, Univ. of Hertfordshire, UK, Sept. 2006.

[3] K. Dunham. Notes on the dance. In V. A. Clark and S. E. Johnson, editors, *Kaiso!: Writings by and about Katherine Dunham*, pages 514 – 519. Univ. of Wisc. Press, Madison, WI, Jan. 2006.

[4] A. Field. *Discovering Statistics Using SPSS (Introducing Statistical Methods)*. Sage Publications Ltd, third edition, Jan. 2009.

[5] T. Flash and N. Hogan. The coordination of arm movements: An experimentally confirmed mathematical model. *J. Neuroscience*, 5:1688–1703, Dec. 1985.

[6] M. Fujita, Y. Kuroki, T. Ishida, and T. Doi. A small humanoid robot SDR-4X for entertainment applications. In *Proc. IEEE/ASME Intl. Conf. on Adv. Intell. Mechat.*, vol. 2, pages 938–943, July 2003.

[7] C. L. Krumhansl and D. L. Schenck. Can dance reflect the structural and expressive qualities of music?: A perceptual experiment on Balanchine's choreography of Mozart's Divertimento No. 15. *Musicae Scientiae*, 1(1):63–85, 1997.

[8] K. MacDorman. Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: an exploration of the uncanny valley. In *ICCS/CogSci-2006 Long Symp.: Toward Social Mechanisms of Android Science*, pages 26–29, Vancouver, Canada, July 2006.

[9] M. P. Michalowski, S. Sabanovic, and H. Kozima. A dancing robot for rhythmic social interaction. In *Proc. 2nd Annual Conf. on Human-Robot Interaction (HRI 2007)*, pages 89–96. March 2007.

[10] C. Reas and B. Fry. Processing.org: programming for artists and designers. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Web graphics*, page 3, Los Angeles, CA, Aug. 2004.

[11] E. Short, J. Hart, M. Vu, and B. Scassellati. No fair!! an interaction with a cheating robot. In *Proc. 5th ACM/IEEE Intl. Conf. on Human-Robot Interaction*, pages 219–226, Osaka, Japan, March 2010.

[12] T. Takeda, Y. Hirata, and K. Kosuge. Dance partner robot cooperative motion generation with adjustable length of dance step stride based on physical interaction. In *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pages 3258–3263, San Diego, CA, Oct.–Nov. 2007.

[13] S. Wairatpanij, H. Patel, G. Cravens, and K. F. MacDorman. Baby steps: A design proposal for more believable motion in an infant-sized android. In *Proc. 23rd Conv. Soc. for the Study of Artificial Intelligence and Simulation of Behaviour*, pages 139–144, Edinburgh, April 2009.

[14] G. Weinberg and B. Blosser. A leader-follower turn-taking model incorporating beat detection in musical human-robot interaction. In *HRI '09: Proc. 4th ACM/IEEE Intl. Conf. on Human Robot Interaction*, pages 227–228, La Jolla, CA, March 2009.