Separation: Short Range Repulsion

Implementation of an automated aesthetic synchronization system for a dance performance

Michael Krzyzaniak[1], Julie Akerly[1 2], Matthew Mosher[1], Muharrem Yildirim[1], Garth Paine[1]
Arizona State University
[1]School of Arts, Media + Engineering
[2]School of Film, Dance and Theatre
{mkrzyzan, julie.akerly, mbmosher, myildirim, garth.paine}@asu.edu

ABSTRACT

This paper describes the implementation of a digital audio / visual feedback system for an extemporaneous dance performance. The system was designed to automatically synchronize aesthetically with the dancers. The performance was premiered at the Slingshot festival in Athens Georgia on March 9, 2013.

1. INTRODUCTION

Dance is often accompanied by other art forms such as music or visual media. Throughout history and across diverse cultures, the dance and accompanying media have typically been synchronized. Not only are they often synchronized in time, but more importantly (and perhaps more frequently), they are synchronized in expression. Advances in electronic technology have made it possible to automate some aspects of this synchronization. Robots have been made to sense and respond to music with movement [12], and music and other media are often manipulated on the fly based on the movement of dancers [6]. This latter arrangement is particularly useful when the dancers are extemporaneously expressing some underlying concept. Although it has become easy and common to use technology to allow dancers to simply manipulate some parameters of sound or other media in real-time, how to automatically synchronize expression in an improvisational setting remains an open question. To what extent can a system translate the underlying meaning of a dance into the language of music or visual media in real-time? This involves, at the very least, first selecting appropriate metrics and sensors to detect the expressive intentions of the dancers, and second designing media that is capable of expressing the same thing. So far, there is no general solution to either part of this problem, although some attempts have been made [3]. It is not difficult to imagine a very good, nearly general solution to the first part (involving perhaps Laban analysis and skeletal tracking), but the second part probably has innumerable idiomatic solutions. Rather than attempting any complete solution here, we present a case-study containing our solution, which implements Paine's framework [7] for applying Dynamic Morphology to sound generation and gestural interfaces.

2. IMPLEMENTATION

Separation: Short Range Repulsion expresses positive and negative societal interpretations of deviant social behaviour [2, 4].

The piece is divided into several sections that explore different points along this gradient. In each section, dancers improvise movement that expresses a particular interpretation of deviance, and the digital multimedia system attempts to measure their behaviour and maintain aesthetic synchrony.

2.1 Ensemble Sections

Illustration 1: synchronized flocking

In several sections, a group of dancers performs some variation on flocking behaviour [8, 9]. Here, 'flocking' refers both to the movement of the dancers through cartesian space (as in 'boids'), and also to gestural imitation. Because the dance is improvised, and there is no established leader overall, the dancers must continually examine each other for behavioural cues in order to maintain synchrony. This arrangement is reminiscent of the the famous Asch conformity experiments [1], in which test subjects are shown to conform with the behaviour of their peers in unfamiliar situations, even when that behaviour is odd or illogical. Sometimes the aim is to maintain synchrony, and at other times the goal is to to deliberately violate it in certain ways, thereby expressing deviance. Often, the moment to moment expression is the result of creative interplay amongst the dancers. For instance, a dancer who has established momentary leadership might abruptly perform unusual maneuvers in an attempt to make one of the others lose synchrony. At other times, a dancer will fall into the others, in a devious violation of the 'separation' rule of flocking, and the others would catch and physically expel her in an attempt to uphold the rule.

The dancers' behaviour was sensed with an overhead near-infrared camera with a wide-angle lens, infrared illumination, and computer vision software. The software tracks each dancer as a blob in cartesian space. Interpreting the blobs as a mass-point system, the software calculates the the dancers' centroid (barycenter), which is interpreted as representing the average behaviour of the flock. The conformity of individual dancers is then assessed by comparing their behaviour against the centroid's. For instance, individual dancers' velocity and distance to the

centroid are compared against the centroid's velocity and the dancers' average distance to the centroid. The software is furthermore capable of determining the orientation of each blob (the direction each dancer is facing), which could have provided additional information about the dancers' conformity, particularly while they are stationary within cartesian space. However, the orientation-tracker was not robust enough to use in this application. Nonetheless, the blob-tracker, as used, is able to measure conformity while the dancers are stationary because it is highly sensitive to changes in the dancers' shape. If, for instance, one dancer moves an arm or leg away from her body, her blob becomes suddenly elongated, which shows up as a sudden burst of speed (because the blob's center moves). If the other dancers do not also elongate their blobs in synchrony, then this dancer's momentary increase in speed is interpreted by the system as devious. This fortuitous feature allows satisfactory measurements of conformity, even when the dancers are not meaningfully translating their positions in cartesian space.

In order to interpret and provide expressive amplification and feedback for the dancers' degree of deviation from the flock, an audio synthesis algorithm was developed to be reminiscent of the two-tone pneumatic sirens (dee-doo, dee-doo) commonly heard on European emergency vehicles. While these sirens are, by context, harbingers of punishment and misfortune, their discrete pitches give them (in the composer's opinion) a more sweet, musical quality than the continuously pitched sirens (wooo, wooo) often heard in the United States. This ambiguity between soothing and alarming makes them well suited to realizing the artistic goal of expressing social deviation as either a creative privilege or criminal activity. By allowing the continuous control of such parameters as the number of contemporaneous sirens, the speed of alternation between the two pitches, the pitch, relative loudness, overall rhythmicity and harmonicity, the amount of doppler shift and other parameters pertaining to a variety of filters, the synthesizer is capable of producing a wide range of overall sounds, that are supposed to range from soothing to alarming, where both qualities are highly multidimensional. Furthermore, in order to highlight the concept of synchrony, the synthesized siren-like sounds were made to be self-similar on two levels of interpretation: the waveform of a given siren is at all times identical to the LFO that modulates its frequency. The waveform is a trapezoidal pulse-train whose slope and duty cycle are under continuous control.

These musical parameters are mapped differently to the metrics of the dancers' performance during each flocking section. A characteristic arrangement consists of several sirens of different pitches mapped to each dancer. The pitches of all the sirens of all the dancers cumulatively form a soothing chord. As an individual dancer's behaviour deviates from that of the centroid, the sirens associated with her become louder, the speed of alternation between pitches becomes faster, and the amount of doppler shift increases (making those pitches out of tune), thereby drawing negative attention to the deviance. Contrastingly, at times the geometry of the waveform and LFO is mapped to a metric of deviance. As the dancers velocity or blob orientation deviates from that of the centroid, the waveform becomes less square and more triangular, which reduces the number of grating higher partials. Similarly, the LFO becomes more triangular, which, in the context of square sirens, sounds whimsical and 'creative'. This casts the deviation in a positive light. During each section only some of the dozen or so available parameters were under control of the dancers. The other variables were either set to constants or were periodically manipulated by the composer during performance. This allowed each section to sound aurally unique and have unique expressive capabilities, as was demanded by the conceptual outline of each section.

2.2 Solo Sections

The overall form of the piece is a rondo. The flocking sections, which explore group social behaviour, alternate with contrasting solo sections that explore individuality and aloneness. A separate audio / visual / sensing system was built for each solo section, which shall be discussed presently.

2.2.1 Internal Combustion

Illustration 2: dancer emerging from the 'introverted' pose

One solo section expresses the increasing frustration associated with extended periods of social isolation, and the eventual self-discovery and strengthening of one's own personality that results. The dancer represents this using a score containing two parts. During the first part, she spends several minutes going slowly from a pin-like standing shape to a ball-like crouching shape, and then returning to standing. Subsequently, during the second part, she dances for several minutes in a highly unrestrained, energetic, spontaneous and 'creative' manner. During the first part of the dance, she literally becomes introverted by physically turning her body slowly in on itself, and then then she re-emerges and, in the second part, makes a passionate and unabashed display of her newfound self. It was observed that her neck is perfectly vertical in the pin-like pose, and again vertical but upside-down during the ball-like pose. Thus, the orientation of her neck with respect to gravity is a good measure of how ball-like, and therefore introverted she is. This is sensed by placing an accelerometer on the back of her neck and measuring the gravitational force on her longitudinal axis. During the second part of the dance, her neck moves spasmodically, along with the rest of her body. This spasmodic motion is also detected with the accelerometer. A three-axis device is employed, and the sum of the absolute values of the partial derivatives of the readings of each axis serves as a good measure of how spasmodic the movement is. In retrospect, it may have been better to use a quaternion to calculate the overall magnitude of acceleration, and a high-pass filter to remove the effects of gravity. Because the rate of change of orientation during the first part of the dance is slow, the partial derivatives of the acceleration are effectively zero, and thus the 'spasmodic' metric can be used concurrently with the 'introversion' metric without significantly interfering. In fact, the high level of volatility of the audio visual feedback system to jerky movement (designed for the spasmodic part of the dance) actually caused the dancer to report that she is able to perform the first part of the score more smoothly with the feedback system than without.

The audio system for this solo consists of a noise with a variable 'ugliness' coefficient. The basic sound is several frequency modulated sine waves that collectively form an A Major triad comprising mostly the note A across many octaves. This sonority is a subtle reference to the opening of Mahler's first symphony, in which the octaves perhaps represent perfect harmoniousness in the universe. As the ugliness coefficient increases, the carrier waves become clipped and go out of tune, the index of modulation moves to non-integer values, the overall loudness increases, and the cutoff of a low-pass filter (applied to the entire sound) increases subtly to make the grating higher partials more audible. The result is quite discordant for high levels of ugliness. The dancer's level of introversion is mapped directly onto the ugliness coefficient, so that as she becomes introverted, the 'ugly' sound expresses greater mental intensity. The spasmodicity in her movement causes the pitches in the A Major chord to detune, some increasing in pitch and others decreasing. The amount of detuning is proportional to the spasmodicity of movement, so that a more spasmodic movement results in a more spasmodic sound, and the sound serves as a reward for the dancer that encourages highly spasmodic movement.

2.2.2 One Idea Solo

Illustration 3: dancer signaling the number one

Another solo section represents a model of behaviour in which overbearing or uncomfortable social situations routinely force an introvert to revert to a single comforting solitary activity, which is typically an addiction [11]. The dancer represents this in a highly abstract way. She chooses one simple gesture, and improvises a dance based entirely on variations on this gesture. The gesture, which involves her signaling the number one with her hand, represents the single, comforting behaviour, and she makes departures from and calm returns to this gesture. The sensing is performed by tracking the dancer's overall speed in cartesian space, as seen from an overhead camera. There may exist more appropriate methods of sensing this dance. For instance, if the basic gesture were known in advance, skeletal-tracking could have been employed to determine her instantaneous degree of similarity between the observed movement and a template of the basic gesture. However, this was, for a variety of reasons, out of scope for the project. Nonetheless, because the basic gesture is always performed while essentially stationary, and the dancer's departures from the basic gesture are always accompanied by bursts of velocity (or changes in blob shape, as discussed above), this functions as a passable metric of the idea being expressed.

The audio system for this section comprises the sound of a crowd with a variable 'screamingness' coefficient. The screaming crowd represents an uncomfortable social situation. As the screamingness coefficient decreases, the crowd goes continuously from screaming to talking to whispering. This is achieved using granular synthesis, where grains are chosen randomly from several audio files that are categorized as containing whispering, talking and occasional laughter, or screaming. Raising the screamingness coefficient makes it more likely that a new grain will be taken from a recording in the the screaming category, whereas lowering it increases the probability that it will be chosen from the whispering category. Increasing the screamingness coefficient also causes the grain density to increase, the shape of the window function to become more square, and the grain duration to become shorter, resulting in a denser, choppier, more 'stressful' texture. Additionally, an *a capella* Armenian song is played over the crowd noises. As the screamingness coefficient increases, the song's playback becomes more fragmented (it is windowed by a pulse train whose duty-cycle decreases with increasing screamingness coefficient) and quieter, as though it is being drowned out by the crowd's screaming. The dancer's velocity is mapped directly to the screamingness coefficient. This allows the social discomfort expressed by the sound to remain synchronized with the same as expressed by the dance.

2.2.3 Hair Dance

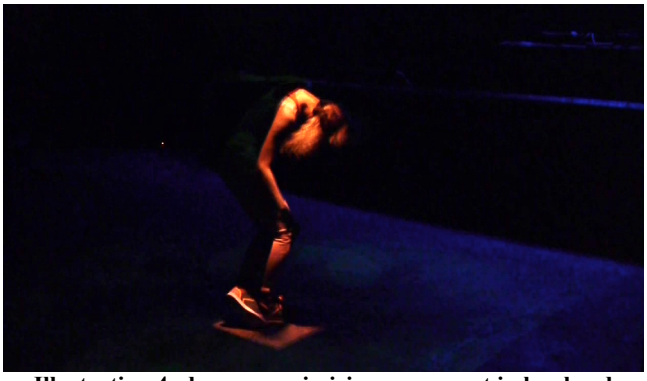

Illustration 4: dancer maximizing movement in her head

Yet another solo section represents an unhealthy separation of mind and body, as occurs, for instance, in depersonalization disorder [10], and the resulting loss of connection with society and the rest of the world. The dancer represents this separation by deliberately violating the patterns of total body connectivity described by the Bartenieff movement fundamentals [5]. Specifically, she violates the homologous pattern which mandates that the upper and lower body function as an integrated unit. Initially, she was going to perform a dance in which she moved only her body, while attempting to keep her head perfectly still in space, thereby representing the disconnect between mind (head) and body. Ultimately she did the opposite: she performed a dance entirely with her head. The dancer's movement was sensed by placing a three-axis accelerometer on the back of her head, and measuring the sum of the absolute values of the partial derivatives of the readings of each axis, which served as an indication of the general level of motion of her head. It would have perhaps been more appropriate to use two three-axis gyroscopes, one on her head and a second on her torso, and measure the difference between each axis of the two, since the aim of the dance is in fact to maximize this difference. With appropriate audio / visual feedback, this arrangement could have encouraged a highly refined and nuanced representation of the concept. Contrastingly, the sensor and feedback, as used, only encourages her to

maximize acceleration in her head, which to some degree, in practice, requires the coincident use of her entire torso and legs.

The audio system comprises crashing and banging noises which are fun to produce but somewhat jolting to hear. This was intended to encourage the dancer to move her head (separately from her body), while still representing the disjointed nature of the underlying expressive concept. These noises are recordings of percussive sounds played back at varying speeds, such as to constrain the pitches to a scale. Two categories of recording are used: moderately jarring sounds taken from a ceramic drum, and highly jarring sounds taken from a hammer striking a metal sheet on an anvil. Whenever the amount of movement of the dancer's head exceeds a certain threshold, a moderately jarring noise is selected at random and played back at a random speed (within the pitch constraint). When the level of movement exceeds a higher threshold, a noise is selected from the highly jarring category. Her sensor is polled at irregular intervals centering around 50 ms. Every new sensor reading above the appropriate threshold triggers a new recording, so that an abrupt movement of the head is accompanied by a battery of crashing and banging. Even if better sensing had been used, this sound was probably not nuanced enough in timbre to provide the dancer with highly detailed feedback of an appropriate nature, because its only expressive control was discrete rather than continuous. Perhaps a sound with two components, with a continuously varying level of dissimilarity would have been more appropriate. Nonetheless, ignoring these conceptual considerations, the audio was successful in that the dancers felt it to be the most responsive part of the piece, and as a viewer it felt like the sound represented the movement (if not the underlying concept) well.

3. EVALUATION

The group-flocking system was informally tested with users during the development phase. At that time, the audio / visual system was configured to jump dramatically and discontinuously from very calm to very alarming when the flock size exceeded a certain threshold. Upon entering the space, the participants, who knew flocking was somehow involved, began walking around in a group. Individuals quickly realized that they could trigger the change by moving a few feet away from the group, and began doing so deliberately and playfully. In a sense, this was the opposite of the intended behaviour, because the system's change from calm to alarming was intended to discourage people from leaving the group and to punish them when they did. In another sense, however, their behaviour was more complex. Had the group disbanded entirely, with every individual choosing to leave, the feedback system would have remained in its alarming state because the flock's radius would never have become small enough to trigger the calm state. Therefore, the satisfaction derived from the devious behaviour of individuals depended on the conformity of the rest of the group. So, what appeared to happen was that the group tacitly agreed to maintain a baseline level of conformity, while allowing individuals to take turns making expressive deviations. This is consistent with the original goal of making a system that comments upon the fine distinction between positive and negative interpretations of deviant social behaviour.

4. CONCLUSION AND FUTURE WORK

In general, the claim that a system maintains aesthetic synchrony with a dancer's intention in real time requires more rigorous evaluation than presented here. Such claims should be cross-validated against the dancer's actual experience. At the time of writing of this paper, the authors are conducting a phenomenological study that attempts to address this. Several dancers are observed interacting with different parts of the system in isolation, and subsequently queried about their subjective experience. The results of this study are forthcoming.

5. ACKNOWLEDGMENTS

Thank you to Arizona State University's School of Arts, Media + Engineering, School of Film, Dance and Theatre, and Graduate and Professional Student Organization for travel funding, and to the University of Georgia's Ideas for Creative Exploration for securing a performance venue and providing logistical support. Thanks to Grisha Coleman and Pavan Turaga for mentorship, and to the dancers Denise Stein and Eleanor Hanafin, whose ideas played an important role in the creation of the work.

6. REFERENCES

[1] Asch, S. Opinions and Social Pressure. *Scientific American*, 193:5 (1955), 31–35.

[2] Ben−Yehuda, N. Positive and negative deviance: More fuel for a controversy. *Deviant Behavior,* 11:3 (1990), 221– 243.

[3] Camurri, A., Volpe.,G, De Poli, G., and Leman, M. Communicating Expressiveness and Affect in Multimodal Interactive Systems. *IEEE MultiMedia 12, 1* (Jan 2005), 43–53.

[4] Goode, E. *Deviant Behavior* (6th ed.). Prentice Hall Publishing, Upper Saddle River, NJ., 2000.

[5] Hackney, P. *Making Connections: Total Body Integration through Bartenieff Fundamentals.* Illustrated by Mary Konrad Weeks. Gordon and Breach Pub., Amsterdam, c1998.

[6] Ingalls, T., Rikakis, T., James, J., Qian, G., Olson., L., Guo, F., and Wong, S. A Movement Recognition Engine for the Development of Interactive Multimedia Works. 2004.

[7] Paine, G. Gesture and musical interaction: interactive engagement through dynamic morphology. *Proceedings of the 2004 conference on New interfaces for musical expression,* (2004), 80–86.

[8] Reynolds, C. W. Flocks, Herds, and Schools: A Distributed Behavioral Model. *ACM SIGGRAPH Computer Graphics* 21, 4 (1987), 25–34.

[9] Shaw, E. Schooling Fishes: The School, A Truly Egalitarian From of Organization in which All Members of the Group are Alike in Influence, Offers Substantial Benefits to its Participants. *American Scientist* 66, 2 (March - April 1978), 166-175.

[10] Simeon, D., Kozin, D. S., Segal, K., Lerch, B., Dujour, R. and Giesbrecht.,T. De-constructing depersonalization: Further evidence for symptom clusters. *Psychiatry Research* 157 (January 2008), 303–306.

[11] Wilson, G. T. Cognitive Process in Addiction. *British Journal of Addiction* 82 (1987), 343–353.

[12] Yoshii, K., Nakadai, K., Torii, T., Hasegawa, Y., Tsujino, H., Komatani, K., Ogata, T. and Okuno, H. G. A Biped Robot that Keeps Steps in Time with Musical Beats while Listening to Music with Its Own Ears. *Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems* (San Diego, CA, USA, October 29 - Nov 2, 2007), 1743 – 1750.