CAPTURING GESTURES FOR EXPRESSIVE SOUND CONTROL

Todor Todoroff, Cécile Picard-Limpens, Julien Leroy & Alain Crevoisier

We present different tools that give musicians extended control for live performances. We developed at Numediart a wireless system of light wearable MARG sensors. We also developed tools to align the orientation of each performer's limb with the skeleton from the Microsoft Kinect camera. The Surface Editor is used to easily and intuitively map sensor data to OpenSoundControl or MIDI messages.
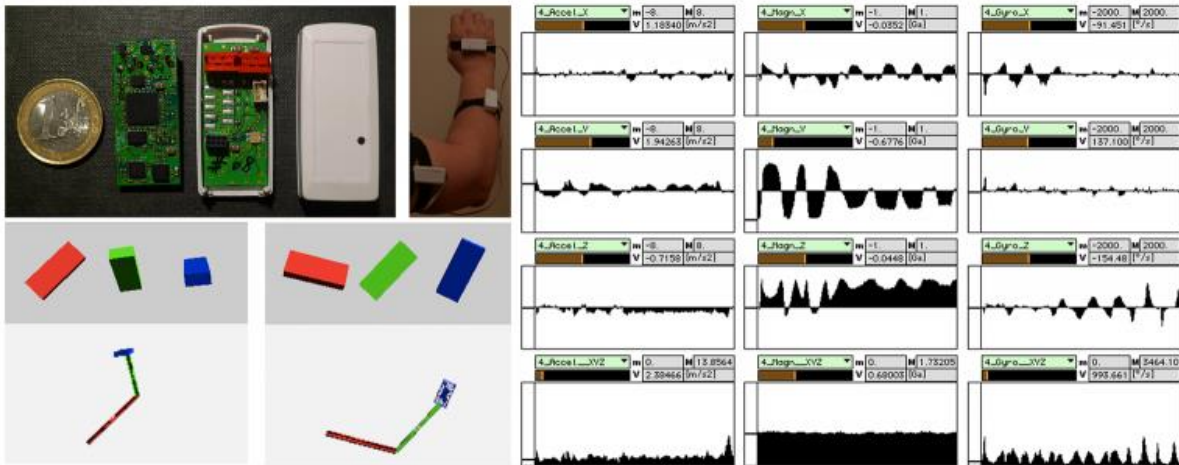
Fig 1. Top left: Sensor node, from left to right: 1 Euro coin, bottom, top PCB views and boxed, with hole for control LED; Bottom left: Attitude of the upper arm (red), lower arm (green) and hand (blue), and reconstructed skeleton in Jitter; Right: Max/MSP Display of the 3 axes of the accelerometer, magnetometer and gyroscope, with total amplitude.
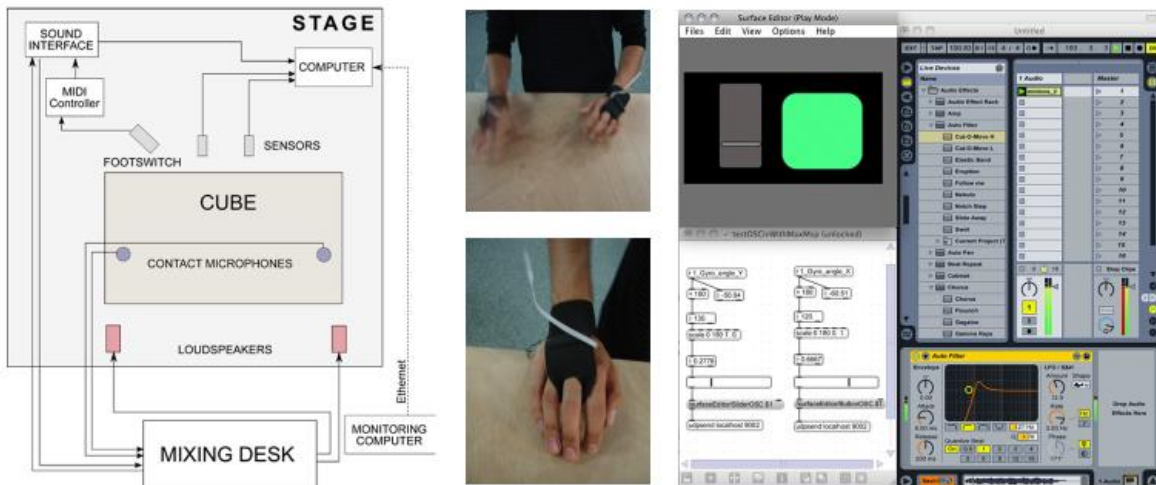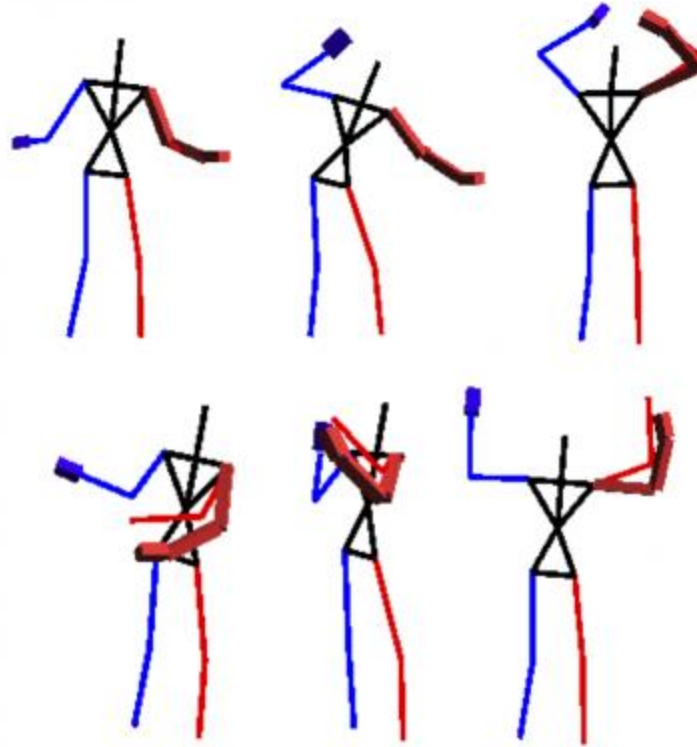

Fig 2. Left: Augmented percussion ("CUBE") with sensors: scheme of set-up; Center: The percussionist wears the sensors on its hands. Gestures orientation are mapped to specific sound modulation parameters; Right: Parameters of the device in Ableton Live applied on Track 1 are controlled by the sensors through the Surface Editor.

Fig 3. Fusion between the Kinect skeleton (thin lines) and the "sensor" limbs (left hand and right arm, solid bodies) in Jitter, aligned in top pictures, while bottom pictures show misalignments due to Kinect latency and its approximative evaluation of elbow and wrist positions.

1. Introduction - Musical Gestures and Interfaces

Musical communication between musicians and listeners, is based on movement: performers control instruments through body movements, which are encoded through audio, and finally analyzed by listeners [Godoy and Leman, 2009]. The concept of gesture, defined as "a movement of part of the body to express an idea or a meaning", and the extensive presentation of its relationships with sound in music have emerged [Godoy and Leman, 2009]. [Cadoz, 1988] proposes a classification of instrumental gestures depending on their function: excitation, modification and selection gestures. New technologies used for musical performance should therefore build meaningful combinations of sound and movement, in order not only to ensure coherence in the musical experience of the performer, but also to preserve the musical communication between musicians and listeners.

We focus here on the idea of "augmented instruments", i.e. acoustic instruments with additional technology capabilities [Miranda and Wanderley, 2006]. It relates to some of the Hyperinstrument projects (http://opera.media.mit.edu/projects.html) at MIT. We aim at extending musical playing techniques digitally in meaningful andintuitive ways while minimizing the performer's constraints. This approach relies heavily on technologies that are able to track the gestures of the performers.

We present in section 2 the MARG (Magnetic, Angular Rate and Gravity) sensors we developed at Numediart [Todoroff, 2011]. The attitude of each sensor can be computed, giving the orientation of each

performer's limb they are attached to. We developed tools to align those orientations with the skeleton obtained with the Microsoft Kinect camera (http://www.xbox.com). It becomes then possible to replace some limbs of the Kinect skeleton by faster and more accurate limbs computed from the sensor data, while keeping the absolute position of the torso given by the Kinect. Our gesture recognition algorithms [Bettens and Todoroff, 2009] show improved performances when used with orientations rather than raw sensor data. The first prototype of these Numediart sensors was used in 2010 in a project with a viola player who dances while playing the viola [Todoroff et al., 2011]. This on-going project explores gestures beyond the usual augmented instrument focus, like leg movements, as it aims to transform the whole body of the performer into a sound body, extending the traditional sound body of the acoustic instruments to the combination performer + instrument. In section 3, we describe the Surface Editor, a mapping tool originally developed to create control interfaces with a tactile surface. We used the Surface Editor for interfacing the sensor data to sound devices through MIDI or OSC (http://opensoundcontrol.org/) protocols. We finally present a project at HEM intended to focus on the percussionist's movements, and, in particular, those movements that do not directly produce sounds but are performed when preparing or exiting a sound producing gesture (section 4).

A well-formatted pdf version of this paper can be downloaded at http://www.numediart.org/files/capturing_gestures_for_expressive_sound_control.pdf

2. Tracking and recognizing gestures

2.1. Numediart sensors

We started in 2009 at Numediart the design of small wearable sensor nodes that include 3-axial accelerometers, gyroscopes and magnetometers. Our latest design, in 2010, offers 6 additional analog inputs to connect optional sensors (pressure, flexion, light, ...). The 17x38 mm circuit board fits into a tiny USB key plastic box and weights only 5 grams, box included. The nodes may be used as such with a wired USB interface, using a cheap serial to USB bridge. But they were designed to be connected, through a digital I2C Bus, to a dedicated Master node/WiFi transmitter that includes a LiPo battery and a charger in a small 70x55x18mm box, offering about 10 hours autonomy. Up to 8 sensors may be connected at 100Hz sampling rate. More details about the sensors specifications and the Master/Slave node architecture can be found in [Todoroff, 2011].

2.2. Software tools

Sensor data, received either as serial bytes over USB or as UDP packets over WiFi, is decoded by a custom-made Max external that outputs values in meaningful units: acceleration in g, rotational speed in deg/s and magnetic field in Gauss. These values can easily be mapped, within the Max/MSP environment, to sound attributes. And additional processing may be done to extract other features, like hit detection from the accelerometer data.
Having MARG sensors, we use the [Madgwick et al., 2010] method to compute the attitude in quaternions. Attitude is the absolute orientation in the 3D space defined by the earth gravity and magnetic fields. We offer the user a two step procedure: a calibration step, facing magnetic North, to compensate for sensors misalignments on the body, followed by a rotation around the vertical axis, to define the chosen direction for the performance. Quaternions can then be transformed to equivalent Euler sequences of rotations and those angles may be mapped to sound parameters.

If enough sensors are placed on a limb, quaternions can be used to animate a skeleton, giving the relative positions of joints. We may also easily compute the angle between any two given segments, like the angle of the elbow, a very useful feature to map to sound.

2.3. Fusing Numediart sensors and Kinect skeleton

Knowing its position, the Kinect from Microsoft (http://www.xbox.com) is able to reconstruct a 3D scene that provides an absolute reference to the real world. Our application, built on OpenNI libraries and drivers, detects users, tracks their skeleton, transforms camera-centered coordinates to stage-related coordinates, and sends the individual joint positions as OSC messages. The Kinect tracking lacks precision as some joints, like wrists or ankles, are not detected. The skeleton therefore doesn't provide individual segments for lower arm and hand, or lower leg and foot, but "virtual" segments that combine both.

Connecting limbs from our sensors to the Kinect skeleton, we keep the good absolute position of the torso and shoulders from the Kinect, as well as a useful approximation of limbs not equipped with sensors. And we track, with sensors, body parts (like hands or feet) or rotations of limbs around their own axis (like arm or wrist twists) that cannot be tracked with the Kinect. With their lower latency and higher 100Hz sampling frequency, sensors follow more accurately those body parts that, because of their lower inertia, are able to move faster. While accelerometer and gyroscope data may still be mapped directly to sound processes, independently of the attitude and skeleton.

2.4. Gesture recognition

We presented in [Bettens and Todoroff, 2009] a multigrid implementation of Dynamic Time Warping (DTW), adapted to sensor data and available as Max/MSP/Jitter (http://cycling74.com/products/maxmspjitter/) external objects and patches. It allowed to recognize gestures, from a user bank of pre-recorded reference gestures, without prior segmentation, i. e. on the fly, not knowing when a gesture starts or ends. We implemented new distance estimators directly from the quaternions defining the attitude in 3D space. It has many advantages over the estimation of distances from acceleration and gyroscopic data as we did previously:

- attitude does not depend on the speed of execution of a gesture, giving consistent distances at all speeds;
- the orientation in the horizontal plane allows better discrimination between similar gestures;
- as attitude data varies slower that raw data, the computation may be downsampled by a significant factor without loosing useful information, reducing the processing load by the same factor;
- distance estimation from attitude data is more efficient, as only one distance, between the reference and the incoming quaternions, needs to be computed;

We recorded for instance one reference gesture for each letter of the alphabet, using 3 sensors. We could then write a text by drawing each individual letters in the air, with a downsampling factor of 8, with hardly no false positives or negatives. We found this test conclusive, as discriminating all the letters of the alphabet is a more complex task than recognizing the reference gestures that would usually be defined for a musical performance.

3. Mapping Gesture Data to Sound Effects

The Surface Editor (http://www.surface-editor.net/) has been developed as a flexible mapping tool [Kellum and Crevoisier, 2009]. It enables users to create interfaces between inputs, e.g. gestures, and outputs, e.g. sound attributes, by configuring components (zones, buttons, sliders, etc.) and attaching actions to them. Those actions will be processed when specific user-defined conditions are met: for instance, an action attached to a slider can be triggered continuously or only when the slider value changes. Originally conceived for a tactile interface, the Surface Editor has been extended to support the input of any device sending OpenSoundControl (OSC) information. In that way, the Surface Editor is able to gather different input variables from external hardware controllers, such as sensors, in a coherent manner. In addition, it is possible for one controller to change the behavior of another one.

We set the communication from our sensors to Ableton Live (http://www.ableton.com/) using the Surface Editor. Signals from the sensors are sent from the Max/Msp environment via OSC as input parameters to the Surface Editor. The Surface Editor supports LiveOSC (http://liine.net/livecontrol/ableton-liveapi/liveosc/), allowing Ableton Live to inform it of all the available destinations: volume, clips (audio samples) and devices with all their parameters. The user can then map a sensor input action to the desired Live destination simply by selecting it from a dropdown menu. This greatly simplifies the mapping workflow.

4. Experiment with students at HEM

4.1. Approach

We distinguish between sound effects that concern the sound source and those that are related to the sound propagation. In the case of the percussion instrument, the attributes of the instrument (geometry, size and material), the position of the hit on the surface and the characteristics of the excitation strike (i.e., the 'rigidity' of the striking finger(s)) affect the sound production. For example, we can extend the sound production possibilities of the given musical instrument by applying an effect that simulates the acoustic characteristics of other resonant objects. This may be mainly done by acting on the frequencies, amplitudes and decay rates of the resonance modes of the instrument. On the other hand, the sound propagation is mainly depending on the acoustic environment, i.e., the space configuration and the potential secondary sources that affect the amount of reverberation. These elements are rather fixed during a 'classical' performance, but we can add sound effects that simulates a virtual acoustic space where sounds appear to originate from a specific direction in space.
The sensors allows to detect static data such as the angle of the hand in the three directions, but also dynamic data such as the amount of energy of a movement (deduced from the output of the accelerometer). These elements can be used as gesture attributes for mapping sound effects. Having a sensor on each hand can allow enabling/disabling one sound effect with one hand, whereas the other act on modulation parameters. Also, sound effects including several parameters can be better controlled by the use of two hands, such as the reverberation effect where the amount of delay and filtering can be treated separately. If we focus on the percussionist's movements, there are several gestures that do not directly produce sounds and can be used for sound modulation. When exiting a striking movement, the hand movement of the percussionist is progressively slowed down. This movement decrease can be easily used to modulate the amount of an effect. On the other hand, the preparing gesture, e.g., the impetus of a strike, can be handled to set a specific sound layer before the effective strike sound.

4.2. Performance

We collaborated with two students of the Haute Ecole de Musique de Genève (HEM), one in the composition class and the second in the percussion class, both with a contemporary music aesthetic. Our purpose was to provide a technology that can not only be easily integrated by the instrumentalist, but that also allows new trends and ways of composition. The development and configuration of this new musical tool was the result of a constant dialogue with the composer and the percussionist, providing an added musical dimension, straightforward for the composer and non-intrusive for the performer.

The sensors were attached to the hands of the percussionist in a way that doesn't hinder his movements, a very important factor for the performance. The composer defined specific ways to notate the new musical gestures in the score. The percussionist was not asked to perform unusual gestures. The composer focused instead on investigating percussionist's gestures that do not sound in reality, and used those to modulate the sound.

The percussion instrument is a simple wooden cube, referred to as the "CUBE". The signals of the contact microphones are used to amplify or modulate sound. A footswitch allows the percussionist to switch between playing modes.

5. Discussion and Conclusion

We presented solutions for musicians to augment musical performances. We proposed a wireless wearable sensor system and a Max external object that delivers the attitude of each body part equipped with a sensor. Sensors don't provide absolute positions. But we showed how data from a Kinect camera can be combined to give the absolute position of the torso and "sensor" limbs attached to it, as well as the approximate position and orientation of limbs not equipped with sensors. The user can record a bank of reference gestures he wishes to recognize and use our multigrid DTW implementation to do so. We then introduced a software tool, the Surface Editor that allows to map sensor data to sound attributes, while preserving the close relation between musical gestures and sound processing.

In the future, we plan to combine the sensors with the Airplane [Crevoisier and Kellum, 2008], a previously introduced device that uses computer vision to track the interaction on a 2D surface, wether with hands, mallets or sticks. This alternative to the Kinect would allow to combine the attitudes of percussionist gestures with precise absolute location of contact points on the performing surface.

6. Acknowledgments

References and Notes:

1. Bettens, F. and Todoroff, T. (2009). Real-time dtw-based gesture recognition external object for max/msp and puredata. In Proc. SMC '09.
2. Cadoz, C. (1988). Instrumental gesture and musical composition. In Proc. of the 14th ICMC.
3. Crevoisier, A. and Kellum, G. (2008). Transforming ordinary surfaces into multi-touch controllers. In Proceedings of NIME-08.
4. Godoy, R. I. and Leman, M. (2009). Why study musical gestures? In Godoy and Leman (Eds.), Musical Gestures: Sound, Movement, and Meaning.
5. Jensenius, A. R., Wanderley, M. M. , Godoy, R. I., and Leman, M. (2009). Musical gestures, concepts and methods in research. In Godoy and Leman (Eds.), Musical Gestures: Sound, Movement, and Meaning.
6. Kellum, G. and Crevoisier, A. (2009). A flexible mapping editor for multi-touch musical instruments. In Proceedings of NIME-09.
7. Madgwick, S., Vaidyanathan, R., and Harrison, A. (2010). An efficient orientation filter for inertial measurement units (imus) and magnetic angular rate and gravity (marg) sensor arrays. Technical report.
8. Miranda, E. R. and Wanderley, M. M. (2006). New Digital Music Instruments: Control and Interaction Beyond the Keyboard.
9. Todoroff, T. (2011). Wireles digital/analog sensors for music and dance performances. In Proc. NIME '11.
10. Todoroff, T., Benmadhkour, R., and Chessini Bose, R. (2011). Multimodal control of music and fire patterns. In Proc. ICMC '11.