« NAMUDA STUDIES » : DOOPLER RADAR-BASED GESTURE RECOGNITION FOR THE CONTROL OF A MUSICAL ROBOT ORCHESTRA

Godfried-Willem Raes
Ghent University College,
School of Arts & Logos
god@logosfoundation.org

ABSTRACT

This paper describes the application layer for gesture recognition using Doppler-based hardware systems, described in full detail in two papers referenced in the bibliography: one on the hardware used [1] and another on the gesture recognition software engine [2]. The hardware implements a fully non-contact gesture acquisition system based on reflected waves from the naked skin. The recognition software is largely based on fuzzy logic for classification of gesture properties.

Being capable of recognizing a defined set of about twelve expressive gestures in a piece of software is of little significance if an application layer fails. Namuda dance technique [3] requires a mutual adaptation of the performer and the software parameters. In order to make the study of Namuda dance possible, we have designed a series of études in which each single gesture prototype can be practised. Since visual feedback to the performer is very problematic in the context of performance, for it greatly hinders freedom of movement and is by nature too slow, we have opted for auditory display. The robot orchestra [4] as we have designed and built it, makes a good platform for such auditory display, particularly since the sounds are not virtual (loudspeakers) but real acoustic sounds emanating from real physical objects. In fact just about any musical instrument can be seen as an example for auditory display as it by its very nature truthfully converts a certain subset of fine motor skills and gestures into sound. The gestures underlying music practice may very well constitute a basis for the embodiment underlying the intelligibility of music. [5] The motor skills and gestures entailed by playing traditional musical instruments are obviously instrumental in nature. They are dictated by the mechanical construction of the instrument. Therefore, as an extension of the body, an instrument can, at most, be a good prosthesis. By removing the necessity of a physical object, the body becomes the instrument. But this in no way removes the need for motor skill and gestural control.

1. NAMUDA ETUDES

The scheme followed for each étude is always the same: starting with the default parameter settings in the software, the performer has to practice each gesture prototype until the recognition is optimum, as can be judged from the response of the robots. The default parameters are not arbitrary, but have been determined as a result of hundreds of measurement sessions with about twenty different subjects, male and female. Each gesture prototype is mapped to a different subset of responding robots. In this respect, the study of Namuda gestures is quite similar to the study of any musical instrument. A certain level of fine motor control has to be developed in the player. Only once that level has been reached can the recognition software be modified by changing the parameters slightly. One would never buy a new and better violin for a child every time it makes a handling and playing mistake. Only once it knows the basics reasonably well should buying a better instrument become an option. Fortunately, in the case of the invisible instrument, we do not have to buy a new instrument but we can improve the software and adapt it to the player. This last possibility opens a whole new perspective for future developments in instrument building.

1.1. Speedup

To study steady accelerating movements, we made a mapping to the x, y and z vectors for the oboe (<Ob> robot), the cornet (<Korn> robot) and the saxophone (<Autosax> robot) respectively. The speedup property strength is mapped to the pitch the instruments will sound whereas the sound level is controlled by the value of the speed parameter. It is a good exercise to try to have the instruments play as long as possible by stretching the time over which the property can remain in a set state. As soon as the property is set in all three vectors, the large stainless steel shells that make up <Llor> will come into play. The sensitivity is set at a much lower level for this to happen than is the case for the vectors taken separately.

1.2. Slowdown

The slowdown property can obviously only be set when the starting gesture is in movement already. Thus the property can never be triggered from rest. The étude maps all three vectors to pitches in the <Piperola> and

<Bourdonola> robots. These are flue pipe organs. The pitch depends on the value of the slowdown, not on the property strength. When the property is set in all three vectors, the lights on the piperola robot will flash.

1.3. Expanding

To trigger this property a movement is required whereby the amount of moving body surface gradually is increased. The property is associated with growth, explosion, enlargement and can be triggered from a standstill. The x-vector is mapped to our <Klung> robot, an automated Indonesian anklung with brass chimes. Pitch selection is mapped to property strength whereas volume is correlated to the value of the property. The y-vector similarly is mapped to <Simba>, a cymbal playing robot. The z-vector is mapped on <Springer>, a somewhat hybrid robot combining shakers, very large springs mounted on resonators as well as a big motor-driven siren. The selection of the elements playing in this robot is mapped to the duration of the property in the z-vector.

When all three vectors are set, the lights on the <Simba> robot will come up.

1.4. Shrinking

Being the other side of the dipole to the previous property, the gestural associations can also be formulated as imploding, diminishing, getting smaller. Clearly the property presupposes movement to start from. The mapping is as follows: x-vector to <Xy>, a quarter tone xylophone robot; y-vector to <Tubi>, a quarter-tone robot made with thick aluminium tubes; z-vector to <Vibi>, an automated vibraphone. For all three vectors, property strength is mapped to the pitch and property value on the sound level. When the property is triggered in all three vectors the lights on <Xy> will flash.

1.5. Steady

In order to trigger this property it is required that the amount of body surface in motion remain constant within the time frame of measurement. The mapping for this property is on our quarter-tone organ robot <Qt>. When the property is set in all three vectors, the blue lights on the robot will come on. In this mapping the amount of body surface remaining constant determines the pitches of the notes. The attack velocity of the notes is controlled – in a rather subtle way – by the property strength.

1.6. Fixspeed

This property is set as soon as the detected speed of a movement stays reasonably constant within a time frame of 500 ms. It is pretty difficult at first to trigger this property more than just accidentally. The reason is not only due to our control, but also in part due to the cosine factors on the Doppler frequency shifts. The latter can be much improved and even cancelled out by keeping the angle of the movement axis and the sight of the vectorial transducer constant. The mapping of all three vectors here is on one of our smallest robots: <Toypi>, an automated toy piano. When all three vectors have the property set, the lights on the little robot come up.

1.7. Collision

Since determination of this gesture property is based on acceleration followed by a sudden stop, it implies a well-defined sudden stop in the gesture. Rebounding movements should be avoided as they can result in false or double triggering. To make the étude convincing, we mapped the output to nothing but percussion, the collision-based instrument family par excellence. The étude should be performed using all possible parts of the body: not only arms and legs, but also the head, the entire torso, the feet, the elbows and even the belly if musculature allows it... The x-vector is mapped to the drums in <Troms> (a set of drums) and the <Snar> robot, an automated snare drum. The y-vector is mapped to the the cowbells that make up the <Vacca> robot. The z-vector is mapped to the set of woodblocks in the <Thunderwood> robot as well as to the thin metal sheets in the <Psch> robot. When collision is detected in all three vectors, the cymbal in the <Troms> robot will play. If collisions are detected but they are below the sensitivity level set in the software, the white lights on the robots will flash.

1.8. Theatrical Collision

As this prototype is defined as a decelerating movement ending in a stop and accelerating again – avoidance of real collision – it tends to be set more easily over relatively larger time spans. In our étude we mapped the gesture to our <Puff> robot, a percussive organ-like instrument tuned in quarter tones.

1.9. Smooth (roundness)

1.10. Edgy

These gesture prototypes can be practised together. The edgy property strength is mapped to the piano notes, played by our player piano robot. The smooth property is mapped to the pitches of our quarter-tone organ robot, <Qt>. The attack velocity is controlled by the momentary value of the vectorial moving body surface. The property is set based on an analysis of the spectrum of the Doppler signals. Edgy movement with many sudden changes causes an overweight of higher partials in the spectrum. The sound volume from <Qt> is made to be a (slow) function of the value of the low side of the power spectrum. It had to react slowly because it steers the compressor on the organ. Those motors, due to their inertia, cannot change speed suddenly. The recognition quality was rated as excellent by all performers we have subjected to this étude so far.

1.11. Jump (airborneness)

The airborne gesture property is set when the performer jumps and no part of the body is at rest. False triggering can occur on large stepping movements though. The larger or higher the jump the stronger the property will be defined. For this étude the x, y and z-vectors are respectively mapped to <Heli> (a helicon robot) <Bono> (a trombone robot) and <So> (the sousaphone robot). When the property is defined for all three vectors, both castanet-playing robots will join in. The sound volume is mapped on the amount of body mass involved whereas the pitch will be proportional to the property strength.

When all three vectors have the airborne property set, certainty is nearly 100%. If only two vectors trigger the property, certainty is still higher than 90%. Most often it is the Z-vector that fails to trigger, which is easily explainable by the fact that the transducer for that vector is suspended above the performer. In order to be certain that the Z-vector also triggers the property, the jump must also include a positional displacement. Needless to add, this étude is quite exhausting to perform.

1.12. Periodic

This is the least well-functioning property in our set and it definitely needs further improvement. Response time is too slow and false triggers do occur regularly. The mapping to the drums in our <Troms> robot allows practice and evaluation. We have also developed software allowing performers to synchronize midi or audio file playback with gestural input. Within strict limitations such as no sudden tempo changes, avoidance of rhythmical complexities (doublings and triplets) it can even be got to work. Further research is being done based on cepstrum analysis.

1.13. Freeze

To practice this 'non-gesture', we applied an inverse mapping, whereby sound will be heard as long as the freeze property is not triggered. The robot used in this mapping is <Bourdonola>, a low register open organ pipes with a string-like sound. The étude is very fundamental as it lets the performers experience the very high sensitivity level of the system.

Parallel to these recognition-based gesture properties, the implementation also offers a full set of very direct mappings of movement parameters on sound output:

- moving body surface: The most intuitive mapping for this parameter seems to be to sound volume or density.
- speed of movement: The most intuitive mapping for this parameter seems to be to pitch.
- spectral shape of the movement: The most intuitive mapping for this complex parameter seems to be to harmony.
- acceleration of the movement: The most intuitive mapping for this parameter seems to be to percussive triggers.

Of course there is nothing mandatory in the way the mappings of gestural prototypes have been laid out in these études. It is pretty easy to devise mappings more suitable for use out of the context of our own robot orchestra. The simplest alternative implementations consist of mappings on any kind of MIDI synth or sampler. However mapping the data from our gesture recognition system to real time audio streams (as we did in our 'Songbook' in 1995, based on human voice sound modulation via gesture) is an even better alternative.

2. EXTENDING THE TIME FRAME

The gesture prototypes practised in the études reflect a gestural microscale in the time domain. Their validity may be as short as 7ms and can for most properties seldom exceed 2 seconds. The only gesture properties that can persist over longer time spans are freeze, periodic, edgy, smooth, fluent and fixspeed. Some can, by their nature, only be defined over very short time intervals: airborne and collision. These gesture prototypes are to be compared to what phonemes are in spoken language, although they already carry more meaning than their linguistic counterparts. Meaning here being understood as embodied meaning.

By following assignment and persistence of the gesture prototypes over longer time spans, say between 500 ms and 5 seconds, it becomes possible to assign expressive meanings to gestural utterances. Here we enter the level of words and short sentences, to continue using linguistic terminology as a metaphor. When we ask a performer to make gentle movements in order to express sympathy, then the statistical frequency of a limited subset of gesture properties will go up significantly. When we ask for aggression, the distribution will be completely different.

It is on this level of time scale that quite a few of our gestural prototypes may be correlated to the classifications of 'effort' set up by Rudolf Laban in his text 'Effort ; Rhythmic Control of Man's activities in Work and Play'. This text was written well before 1950 and only got published as an appendix to the book mentioned in the bibliography of our paper. (Laban, 1980). As far as we can judge, his classification and notation proposals are difficult to hold in a more generalized context of a theory of expressive gesture. Clearly none of Laban's analysis are based on any other objective measurement than visual observation and the dancers' own experience of effort.

It would be an interesting research project to find out how comparable such distributions for a limited set of sentic forms [6] are amongst a large set of different dancers and musicians. Unfortunately this is beyond the scope of our own mission. In part, such research is being done now by our colleague Dr. Jin Hyun Kim.

3. RELATION TO OTHER DANCE PRACTICES

Although as soon as we gained some insight in the potential for dance offered by our technology – in the mid nineteen-seventies – we carried out artistic experiments with dancers trained in classical ballet as well as modern dance, we quickly found out that such an approach to dance was unsuitable to work well with this technology. Classical dance forms concentrate on elegance and – in general – avoid collision and a sense of mass. Position in space and visual aspects are very dominant. Immediately alternative dance practices came into consideration. In the first place butoh dance, an avant-garde dance practice with its roots in Japan, where we also came in contact with it (through Tari Ito). Thus we got in contact with dancers such as Min Tanaka, Tadashi Endo, Emilie De Vlam and later on Lazara Rosell Albear ... This has led to quite a considerable list of collaborative performances. However, butoh is only vaguely defined from a technical dance point of view. Its non-avoidance of roughness, its nakedness [7] and its concentration on bodily expression, leaving out any distracting props and requisites, formed the strongest points of attraction. Only in some forms of contact improvisation did we find other links, but in this dance form we ran into problems with our technology, which is not capable of distinguishing the movements of more than a single moving body at the same time. As far as couple dances go, we have also investigated tango (and the related milonga) quite a bit, in part also because we happen to be a tanguero ourselves. In this type of dance the problem of the two inseparable bodies poses less of a problem since movements are always very well coordinated. Acrobatic tango is of particular interest for the gestures used are very well defined.

Other dance forms we have experimented with include pole dance, flamenco and break dance. For the first dance style, we even developed a special set of microwave radar sensors that can be mounted at the top of the pole. Unfortunately we found out that the professionals in this dance form have little if any affinity with the art forms we are interested in ourselves, which are after all still quite experimental. As for break dance, we are forced to admit that despite the fascination we have for the virtuosity in movement that can be found in the genre, it seems to be strongly bound to a certain age group that we have long since left behind us...

4. INTERACTIVE COMPOSITION AND CHOREOGRAPHY

It will be clear that the mastering of Namuda opens wide perspectives for the development of real time interactive composition with a strong theatrical component. Over the 35 years that we have been developing the system, many hundreds of performances have been staged.

The entire Namuda system including the invisible instrument is open for use by other composers and performers. Scientists interested in research into human gesture are also invited to explore the possibilities of the system. Other composers that have made use of it so far are Kristof Lauwers and Yvan Vander Sanden. The system has also been investigated by Hans Roels, Troy Rogers, Jin Hyun Kim, Dirk Moelants and others. Many applications have been developed on other platforms than our own GMT-programming environment. To facilitate this, we have designed a limited gesture sensor with built-in signal processing using a PIC micro-controller. This 'Picradar' sensor, as we have baptized it, makes use of microwave radar (9.35 GHz) and outputs its data following the midi protocol.

There is still a lot of work left to be done on improvements to the hardware and recognition software as well as in terms of its artistic implementations. An open invitation.

5. ACKNOWLEDGEMENTS

Thanks to my dance and music collaborators who helped me perform the many necessary experiments and measurements using their bodies: Moniek Darge, Angela Rawlings, Helen White, Dominica Eyckmans, Lazara Rosell Albear, Zam Ebale, Marian De Schryver, Nicoletta Brachini, Emilie De Vlam, Tadashi Endo, Nan Ping, Jin Hyun Kim, Marjolijn Zwakman and many others. Thanks also to my collaborator Kristof Lauwers for helping out with code development, data analysis and debugging.

6. ENDNOTES

[1] The complete hardware description of the system can be found at: http://www.logosfoundation.org/ii/Holosound_ii2010.pdf. The background to our experimental instrument-building projects is largely taken from "An Invisible Instrument", Godfried-Willem Raes, 1997: http://logosfoundation.org/g_texts/invisins.html, which deals in greater depth with the philosophy behind the design.

[2] The software gesture recognition layer is decribed in full in this paper: http://www.logosfoundation.org/ii/Namuda_GestureRecognition.pdf

[3] Namuda is a word of our own casting and stands short for 'Naked Music Dance'.

[4] Detailed information on the robot orchestra and the robots constituting it can be found at http://www.logosfoundation.org/instrum_god/manual.html

[5] These matters were discussed in depth in my texts on the invisible instrument: Raes,1993 ,1994 and 1999.

[6] The notion of sentic forms was introduced by Dr. Manfred Clynes, with whom we had many conversations and discussions at the time we visited him in Sydney when we were demonstrating our invisible instrument at the Sydney Conservatory. References to his publications on the subject can be found in the bibliography section of this paper.

[7] We wrote an essay on nakedness some time ago, after realizing that even nowadays there are still people around that seem to have difficulty in coping with this utmost human property... The text can be read at: http://logosfoundation.org/g_texts/naked.html.

7. RÉFÉRENCES

- BECKMANN, Petr & SPIZZICHINO, Andre "The Scattering of Electromagnetic Waves from Rough Surfaces", Pergamon Press, Oxford, 1963
- CLYNES, Manfred "Sentics, the touch of the emotions" , Anchor Books (NY, 1978)
- DROITCOUR, Amy Diane "Non contact measurement of hearth and respiration rates with a single-chip microwave Doppler radar", Stanford PhD thesis, June 2006.
- KRAMER, Gregory (ed.) "Auditory Display", Addison-Wesley Publishing Company, (Reading MA, 1994)
- LABAN, Rudolf "The mastery of movement", Macdonald & Evans Ltd (Plymouth, 1980)
- LEMAN, Marc "Embodied Music Cognition and Mediation Technology", (Cambridge, 2008)
- RAES, Godfried-Willem "Een onzichtbaar muziekinstrument" (Gent, 1993, Ph.D. thesis)
- RAES, Godfried-Willem "An Invisible Instrument (1994): http://logosfoundation.org/g_texts/invisins.html
- RAES, Godfried-Willem "Gesture controlled virtual musical instruments" (Ghent, 1999): http://logosfoundation.org/ii/gesture-instrument.html
- RAES, Godfried-Willem "Quadrada" (Ghent, 2003): http://logosfoundation.org/ii/quadrada.html
- RAES, Godfried-Willem "PicRadar" (Ghent, 2004-2005): http://logosfoundation.org/ii/picradar.html
- RAES, Godfried-Willem "Distance sensing" (Ghent, 2007): http://logosfoundation.org/ii/distance_sensing.html
- RAES, Godfried-Willem "Naked" (Ghent, 2008): http://logosfoundation.org/g_texts/naked.html
- RAES, Godfried-Willem "Microwave Gesture Sensing" (Ghent, 2009):
- http://logosfoundation.org/ii/dopplerFMradar.html
- RAES, Godfried-Willem "Holosound ii2010, a doppler sonar based 3-D gesture measurement system" (Ghent, 2011): http://logosfoundation.org/ii/holosound_ii2010.pdf
- RAES, Godfried-Willem "Namuda Studies" (Ghent, 2012): http://www.logosfoundation.org/scores_gwr/Namuda_Links/namuda_studies.html
- RAES, Godfried-Willem "Namuda gesture recognition for musical practice" (Ghent, 2011): http://www.logosfoundation.org/ii/Namuda_GestureRecognition.pdf
- ROADS, Curtis "The computer music tutorial", (MIT press, 1996)
- ROELS, Jetty(ed.) "Images of corporeality, traces of Butoh in Flanders", VCIT, (Ghent, 2002)
- SEIFERT, Uwe, HYUN KIM, Jin (eds) "Paradoxes of Interactivity", (Bielefeld, 2008)