

Mobile Music Making History and Prospects

October 24, 2007

Georg Essl

georg.essl@telekom.de

Deutsche Telekom Laboratories
Technische Universität Berlin, Germany

What is Mobile Music Making?

What is Mobile Music Making?

Making music with portable instruments?



What is Mobile Music Making?

Making music in a mobile setting?



What is Mobile Music Making?

Making music using mobile electronic devices.



(Casio VL-TONE VL-1 1981)

What is Mobile Music Making?

Being confused and asking Wikipedia

Mobile music

From Wikipedia, the free encyclopedia

This article is about music on mobile devices, for the musical form see aleatoric music.

Mobile music is music, which is downloaded to mobile phones and played by mobile phones.

Truetones, Chaku-Uta, Chaku-Uta Full

[edit]

While ringtones do not include artists voices, truetones, chaku-uta and chaku-uta full are recordings of artists interpretation of music. Distributing them usually requires the agreement of record labels and other owners of artists rights.

See also

[edit]

- Ringtone

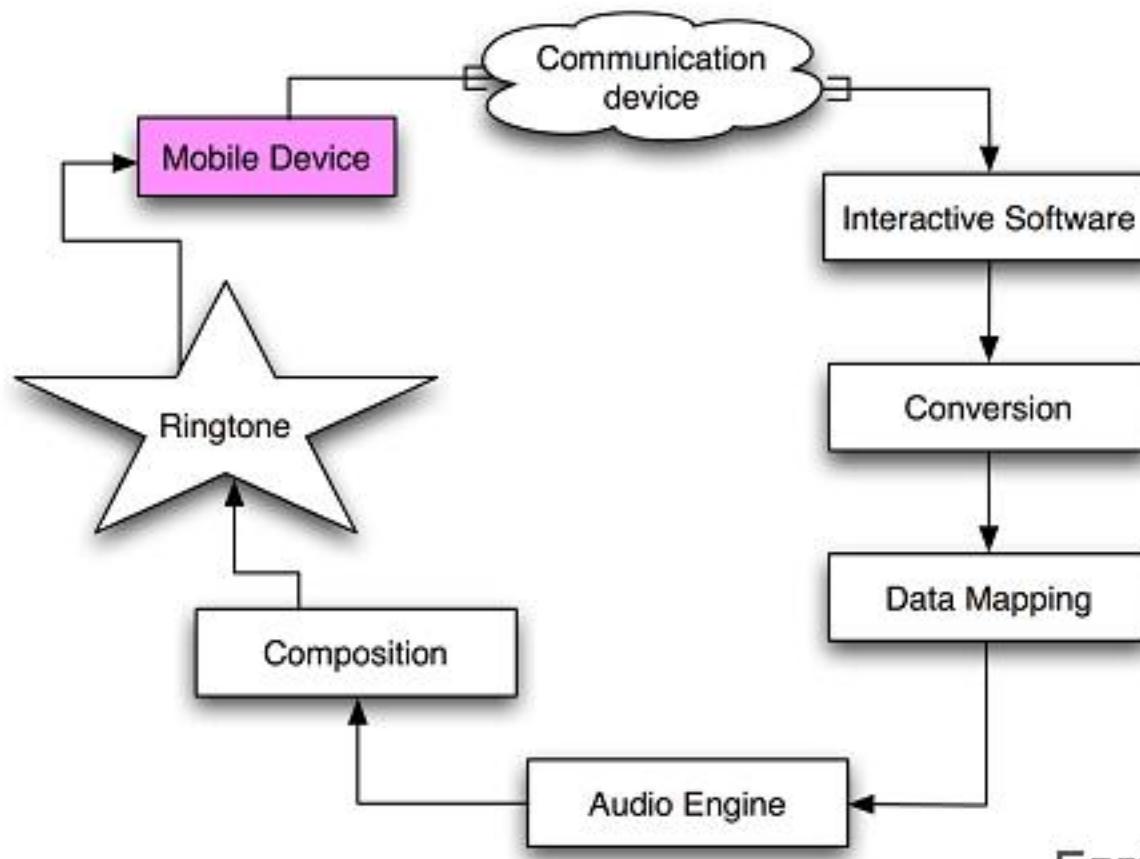
References

[edit]

Category: Mobile phones

Ringtones.

Performative use of Ringtones.



Egotone (Lee 2007)

Mobile Music Making is...

... performing music on general purpose mobile (and mostly commodity) devices.

Mobile Music Making is...

... ~~performing music on general~~
purpose mobile (and mostly
commodity) devices.

Mobile Music Making is...

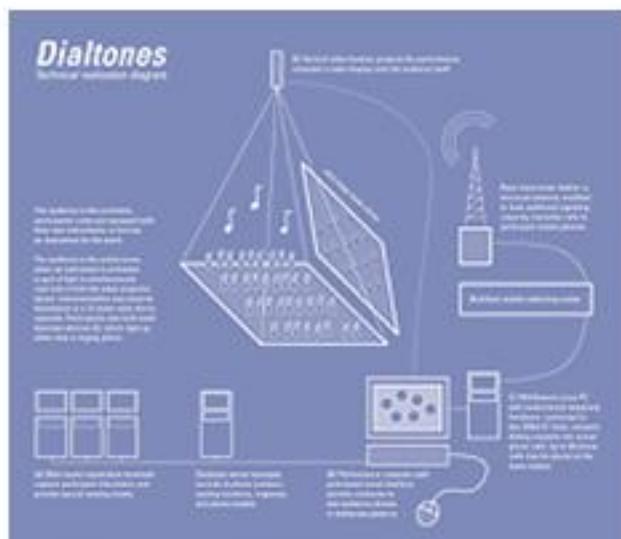
... performing music on general-purpose mobile (and mostly commodity) devices.

What is Mobile Music Making so far?

A Brief History

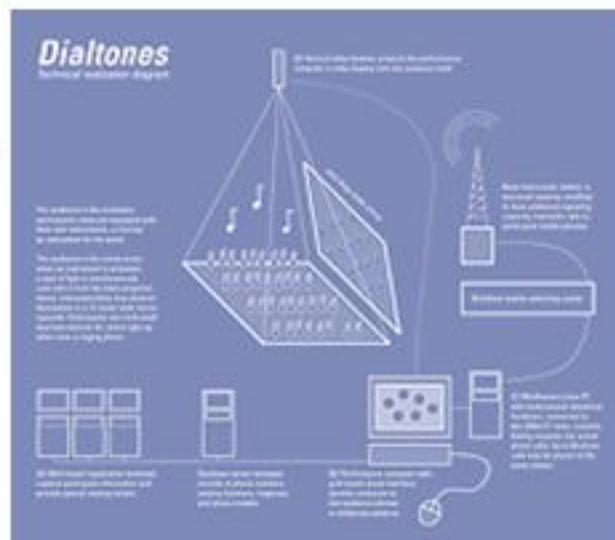
Time Line– Mobile Music Performance

- Passive
 - Wagenaar "Kadoum" (2000)
 - Levin "Dialtones" (2001)
 - Ligna "Wählt die Signale!" (2003)
- Location based
 - Gaye et al "Sonic City" (2003)
 - Carter and Liu "Location33" (2005)
 - Tanaka's net_drive (2006)
- Wearables
 - Maubrey (1982-)
 - Hahn's Pikapika (2001)
- Input-driven
 - Wittchow "Nanoloop" (1998)
- Hybrid physical
 - Schiemer's PocketGamelan (2006)



(Golan Levin 2001)

Mobile Phones in Concert. Levin's DialTones.



DialTones (Levin et al, 2001)

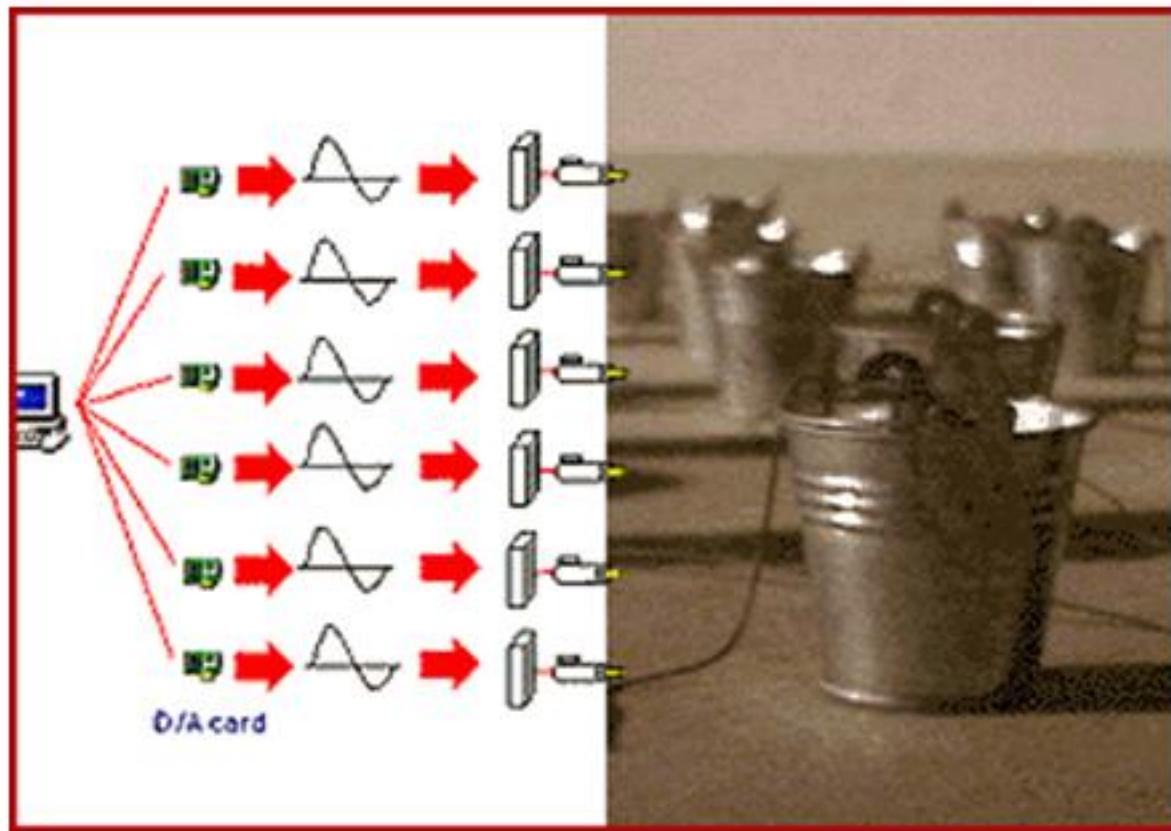
Mobile Phones in Concert. Ligna's "Wählt die Signale".



(Ligna & Röhm 2003-4)

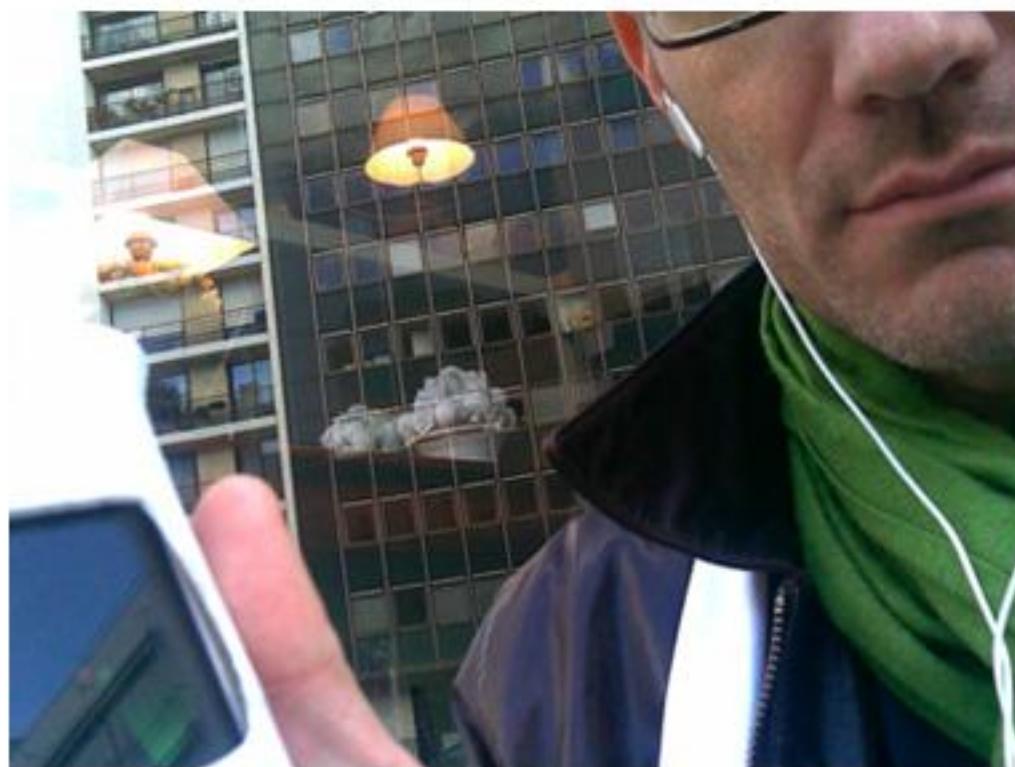
Mobile Networks in Performance.

Wagenaar's Kadoum.



Kadoum (Wagenaar 2000)

Locative Music. Tanaka's "Net_derive".



Net_derive (Tanaka, 2006)

Locative Music. Gaye's Sonic City.



"What I find interesting with mobile music is that it democratises the use of music technology and takes it to the streets."

– L. Gaye 2006 Sonic City (Gaye et al, 2002-2004)

Local Physicality of Mobile Sounding Devices. Schiemer's PocketGamelan.



PocketGamelan (Schiemer, 2006)

Wearable Music. Maubrey and Hahn.



(Hahn et al, 2001)



(Maubrey 1982-2004)

Location-based Music Sharing. Mobile Music Exchange.



tunA (Bassoli et al, 2003)



"Navigation via adaptive music" (Warren 2005)
Location33 (Carter, 2005)
gpsTunes (Strachan et al, 2005)

Gate-based Music Playback.



(Murray-Smith, 2005)

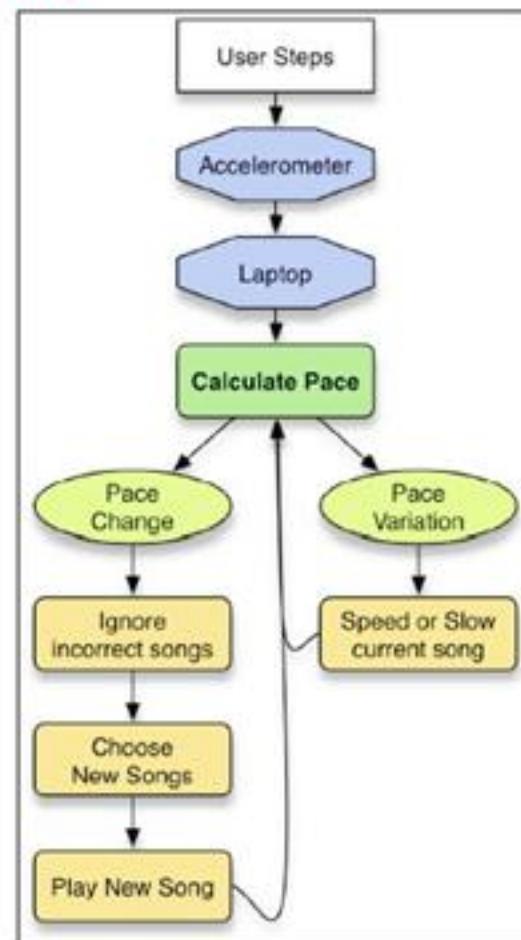
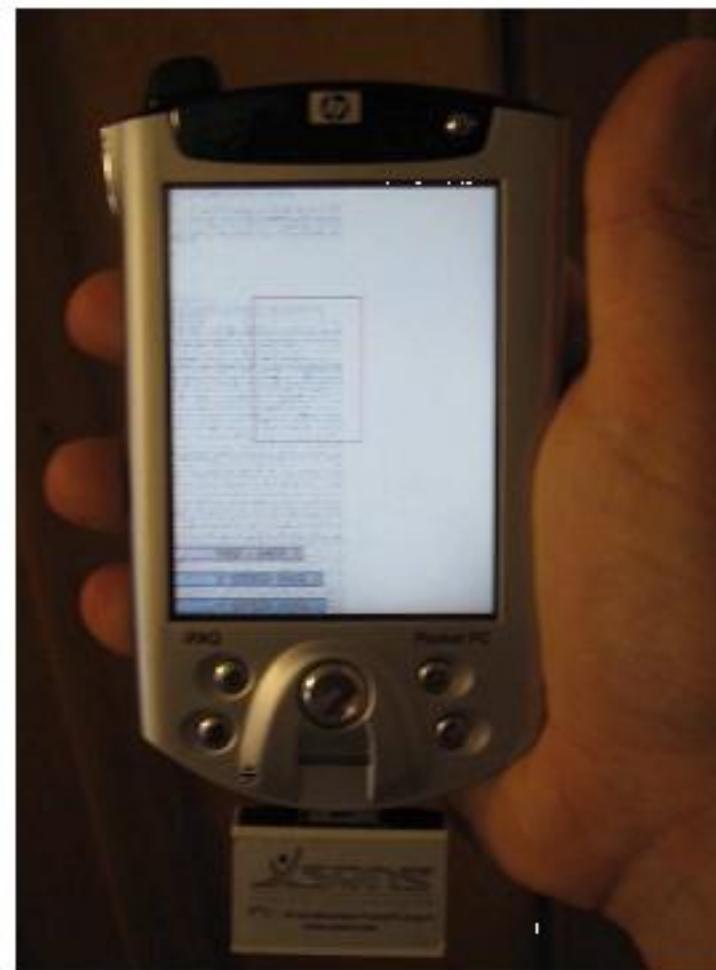
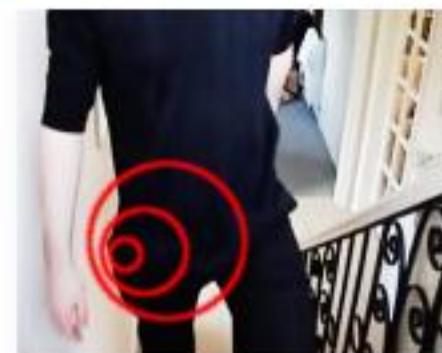
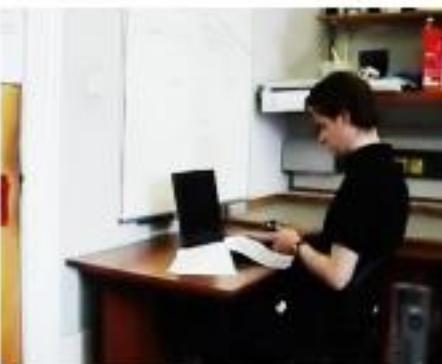


Figure 1. PersonalSoundtrack system diagram showing inputs, actions, and flow of control.

PersonalSoundtrack (Elliott et al, 2006)

Gesture-based Music Interactions. Rhythm detection.



Shoogle (Williamson et al 2007)

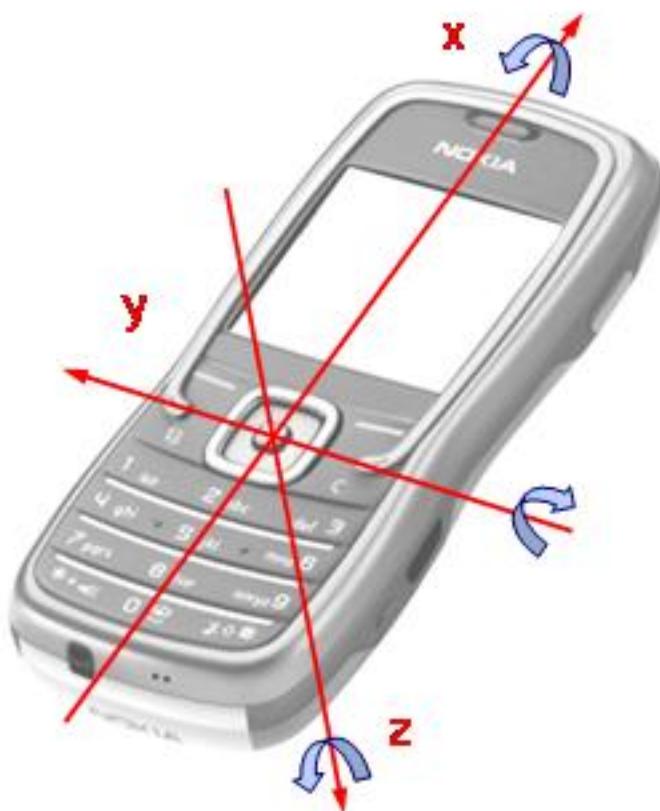
(Murray-Smith et al 2004, 2006)

Gesture-based Mobile Music. CaMus by Rohs, Essl



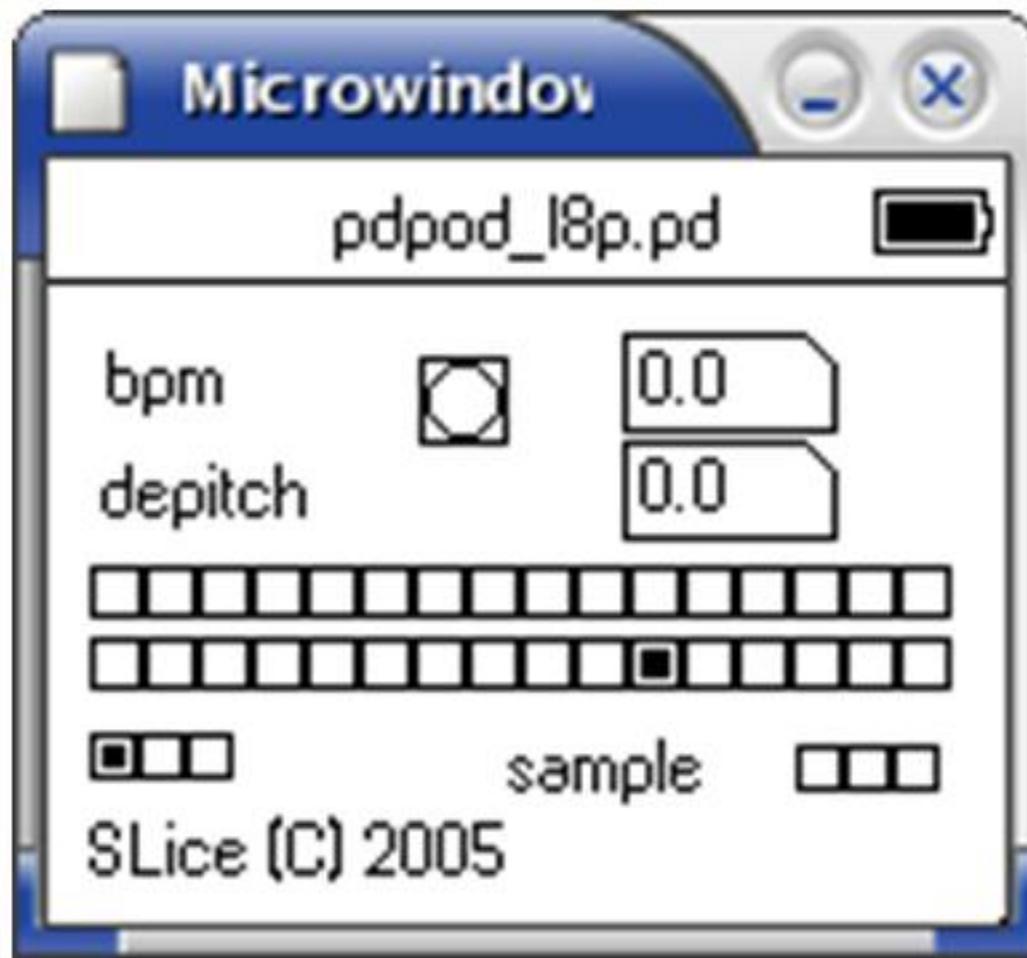
CaMus, CaMus2 (Rohs, Essl 2005-2007)

Gesture Based Musical Interaction. Rohs and Essl's ShaMus.



ShaMus (Essl & Rohs, 2007)

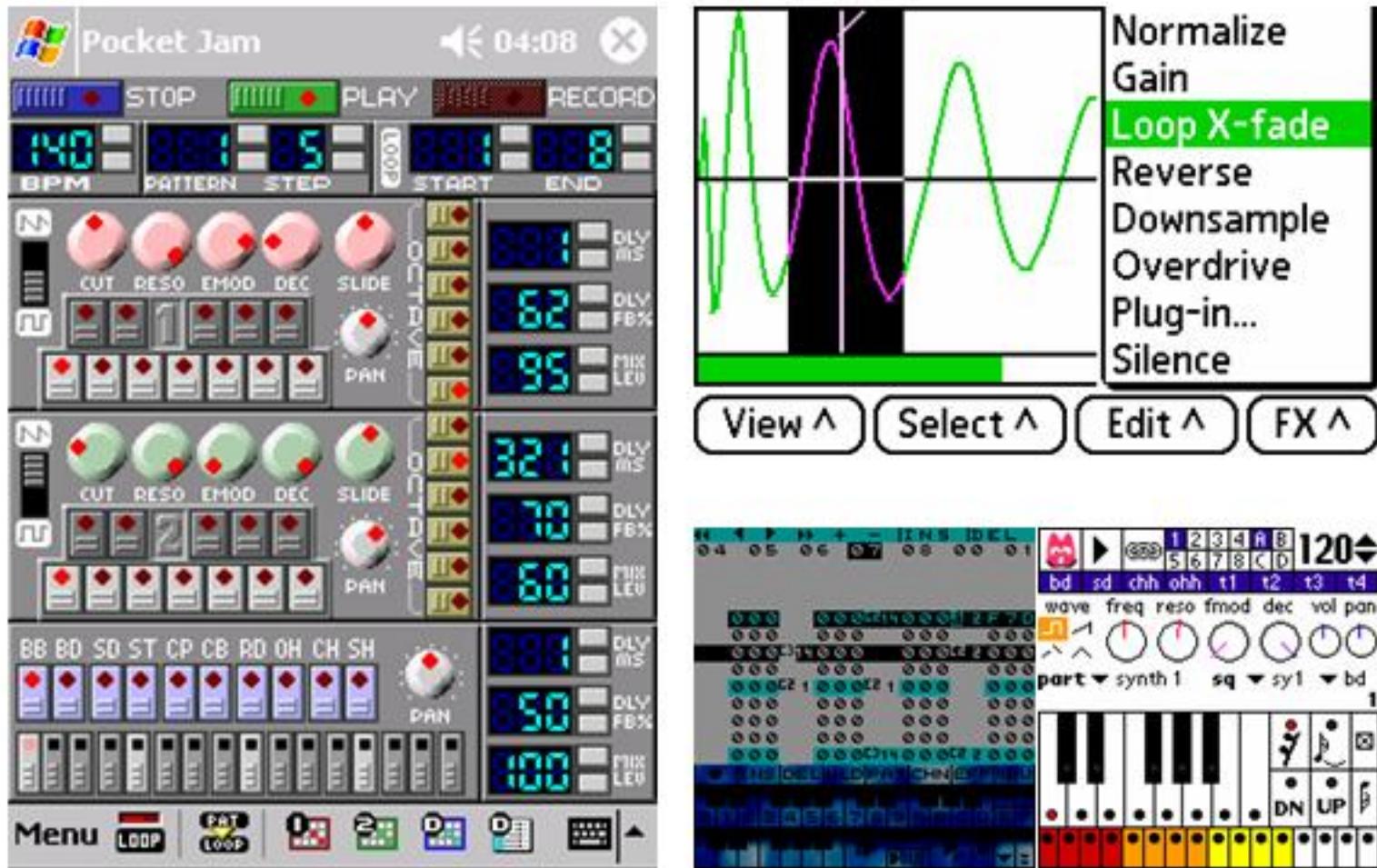
Sound Synthesis on Mobile Devices. PDa and PDpod.



PDa (Geiger, 2003)
PDpod (Kaltenbrunner, 2005)

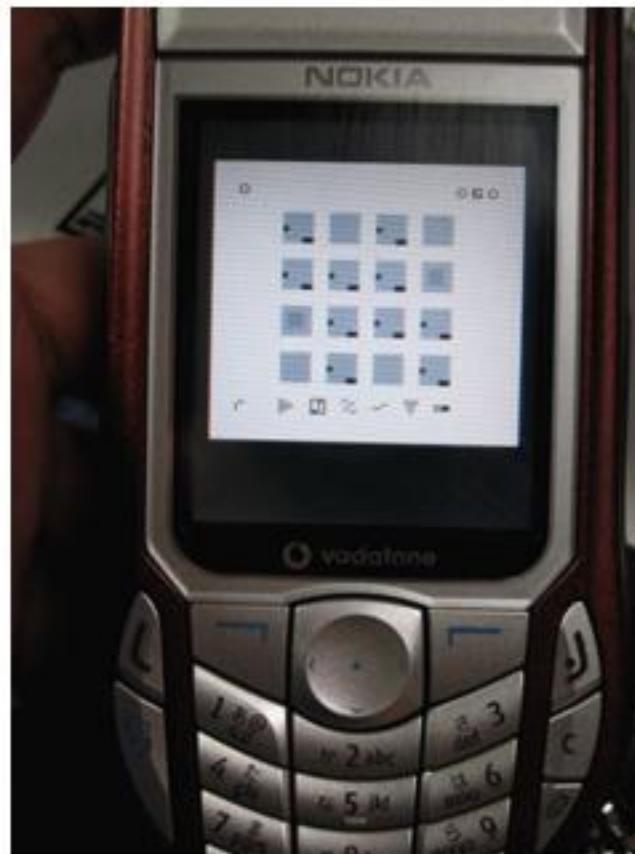
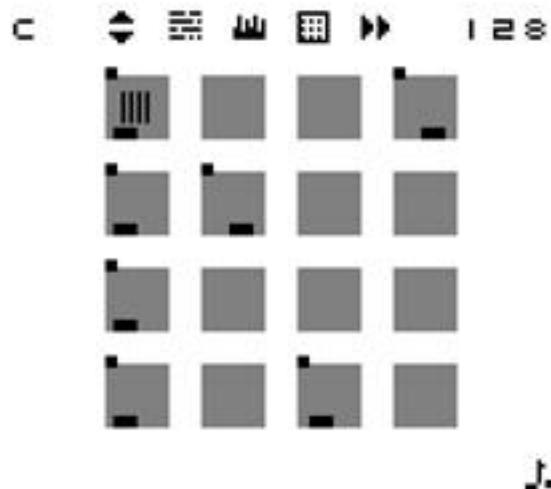
Sound Synthesis on PDAs.

Elsdon Review of PalmOS and PocketPC software



PocketJam, Bahajis Loops, Psytexx, Microbe (Elsdon 2007)

Sound Synthesis on Mobile Gaming Platforms. Wittchow's nanoloop for GameBoy.



Nanoloop (Wittchow, 1998-2007)

Parametric Synthesis on Mobile Phones. Mobile STK.

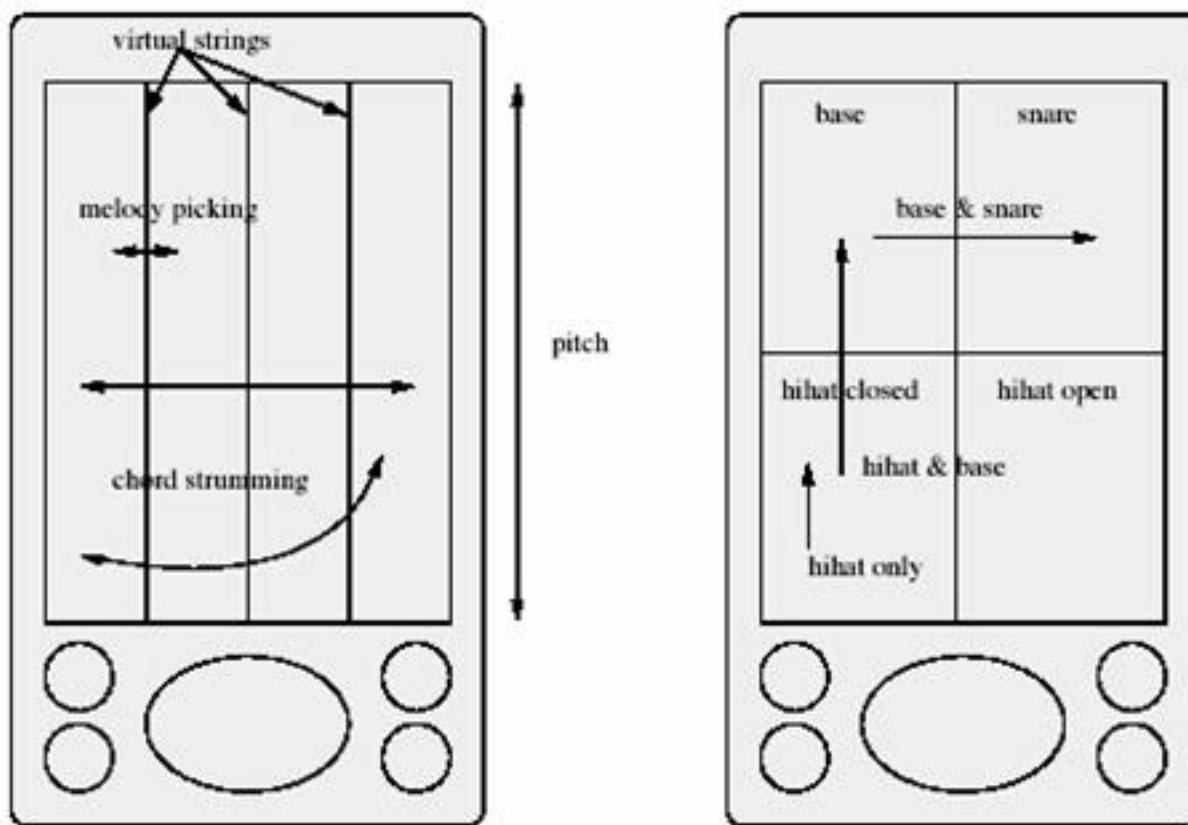


MOBILE STK.SIS

Stk (Cook, Scavone 1997-)

MobileStk (Essl & Rohs, 2006)

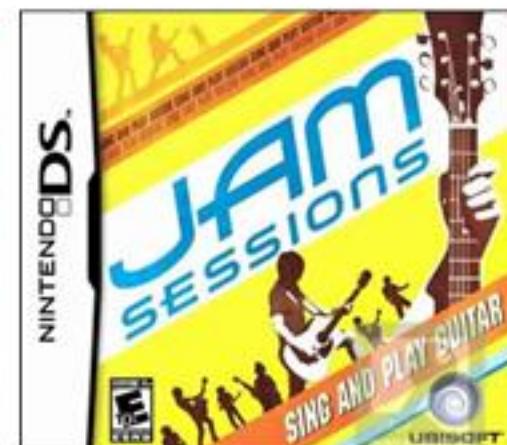
Touch Screen as Input. Geiger's PDa controller.



(Geiger 2006)

Commercial Products.

Nintendo DS/UbiSoft JamSessions.



JamSessions, (Ubisoft 2007)

Mobile Music Making, what next?

Prospects and Challenges

It's the Technology. Challenges and prospects.

Main architecture of mobile phones

- Reasonably good fixed point performance (740 MIPS @ >500 MHz for ARM11)
- No floating point support (very slow emulation, emerging in next generation ARM architecture NEON, vector FPU for graphics)
- Plenty of memory (GBs)
- Audio quality and audio OS layer still changing
- Audio delay changing and still generally too large (.5 secs @ 8000 Hz Nokia 5500)

It's the Sensor Technology. Challenges and prospects.

Available Sensors

- Keys (Many)
- Microphone (Misra)
- Touch screen/stencil (Geiger, JamSessions)
- Accelerometer (Murray-Smith, Essl & Rohs)
- GPS (Tanaka, Murray-Smith)
- Camera (CaMus)
- Magnetometers (ShaMus)
- Gyroscope

Available Actuators

- Screen
- Speaker
- Vibrotactile display

Additional interfacing

- Local networking (Bluetooth/Zigbee)
- Global networking (Mobile phone network, TCP)

Challenges

- Size
- Future availability
- Fidelity
- Commodity

It's the Place.

Challenges and prospects.

Performance types

- Playing alone ("playing on the bus")
 - Local
 - Untethered, power-independent
- Playing portable (mobile camp fire music)
 - Seamless integration
 - Volume, plug&play
 - Ad hoc networking
 - Ad hoc sharing
 - On the fly performance and authoring

It's the Place.

Challenges and prospects.

Performance types

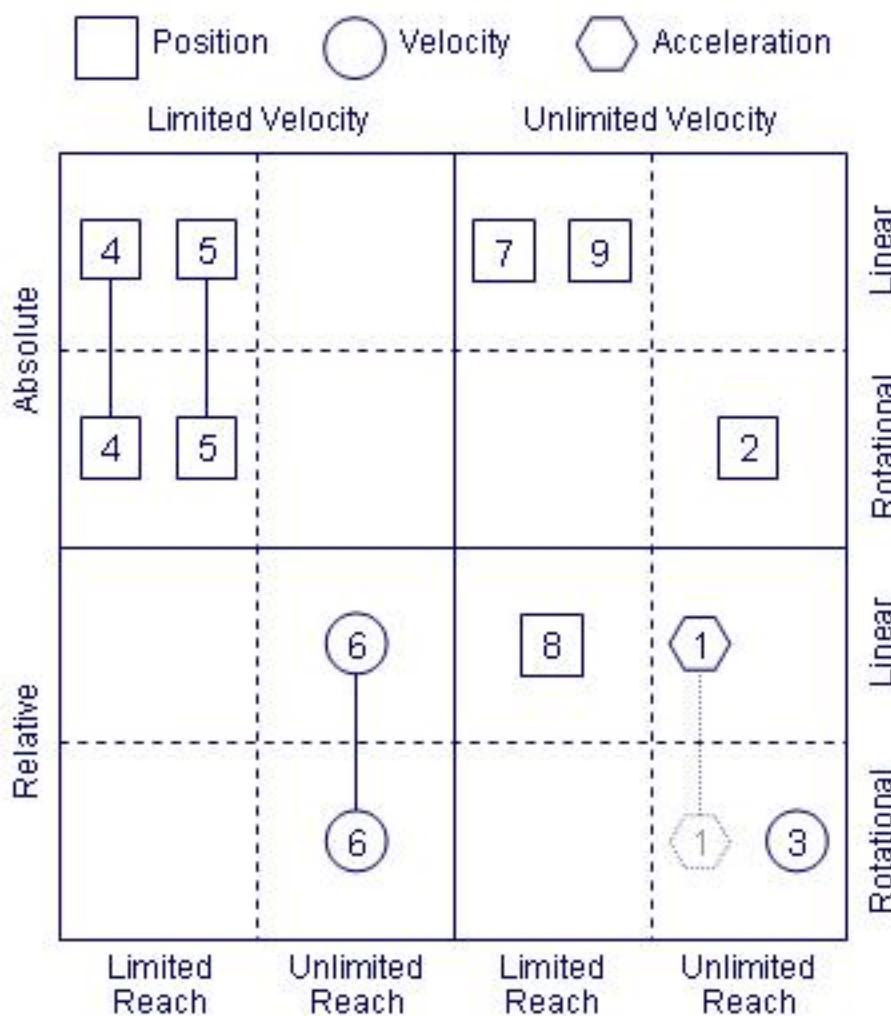
- Playing in ensemble (mobile chamber music, mobile marching band, mobile orchestra)
 - Support pre-meditation (composition, scoring)
 - Support synchronization
 - Support data sharing
 - Support amplification and mixing
- Playing remotely (mobile internet performance)
 - Latency
 - Connectivity
 - Causation

It's the Style.

Composition versus Improvisation.

- **Composition**
 - Pre-meditated intend
 - “Scored”
 - Can be interpreted
 - Scoring requires fixed reference and persistence
- **Improvisation**
 - Flowing intend
 - Largely not scored
 - Requires expressivity

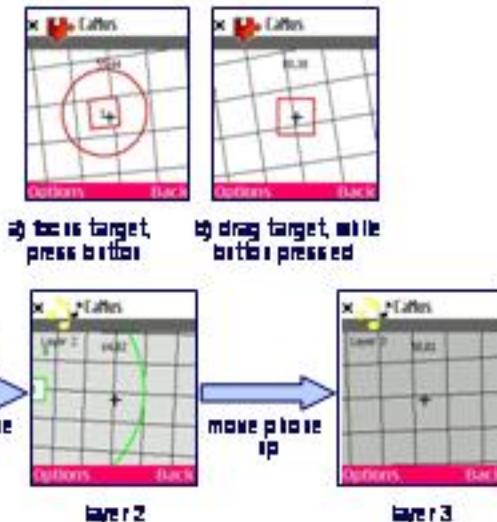
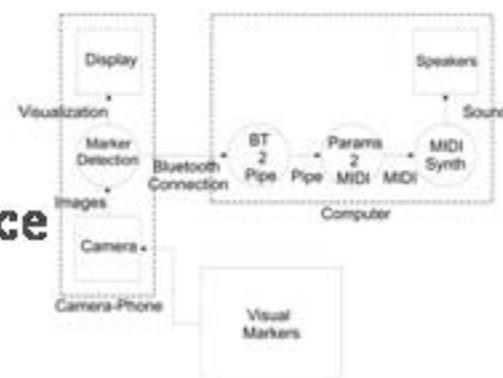
Design Space of Sensors for Mobile Music Performance



1. Accelerometer [m/s²]
2. Magnetometer [Gauss]
3. Gyroscope [degree/s]
4. Visual marker tracking
5. Visual grid tracking
6. Visual movement detection
7. Touch screen
8. Touch pad
9. Capacitive proximity sensor

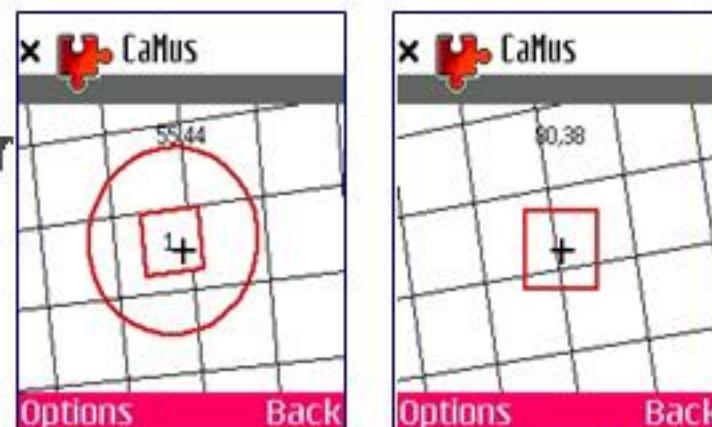
CaMus: Camera Phone Music Performance

- Mobile interactive real time music
 - on-the-fly modification of sequenced music
- Mapping motion to music
 - visual grid tracking, optical flow
 - one-handed interaction by position and distance
 - multiple parameter interaction in parallel
- Visualization on small displays
 - sounding and sound manipulation elements stationary
 - composition via selection and placement of elements
 - interpretation via movement through the space



Dragging and Fixing Targets

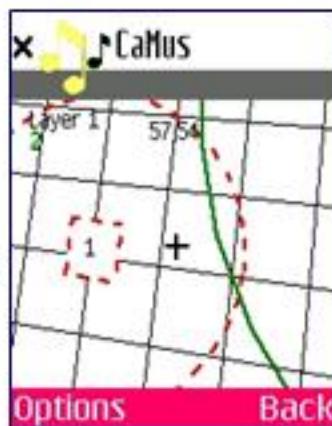
- Creation, deletion, dragging of targets



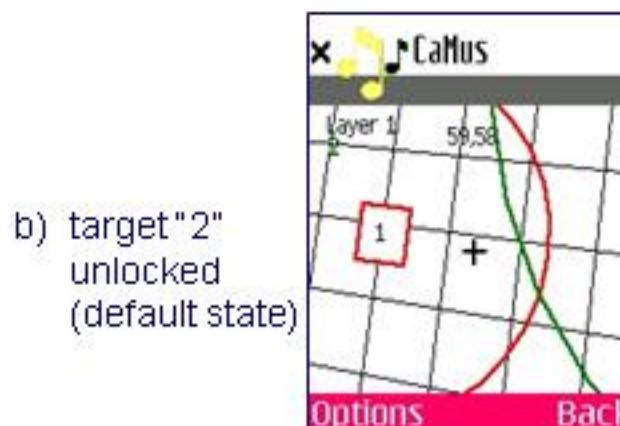
a) focus target,
press button

b) drag target,
while button
pressed

- Fixing targets relative to the cursor (instead of tied to the grid)



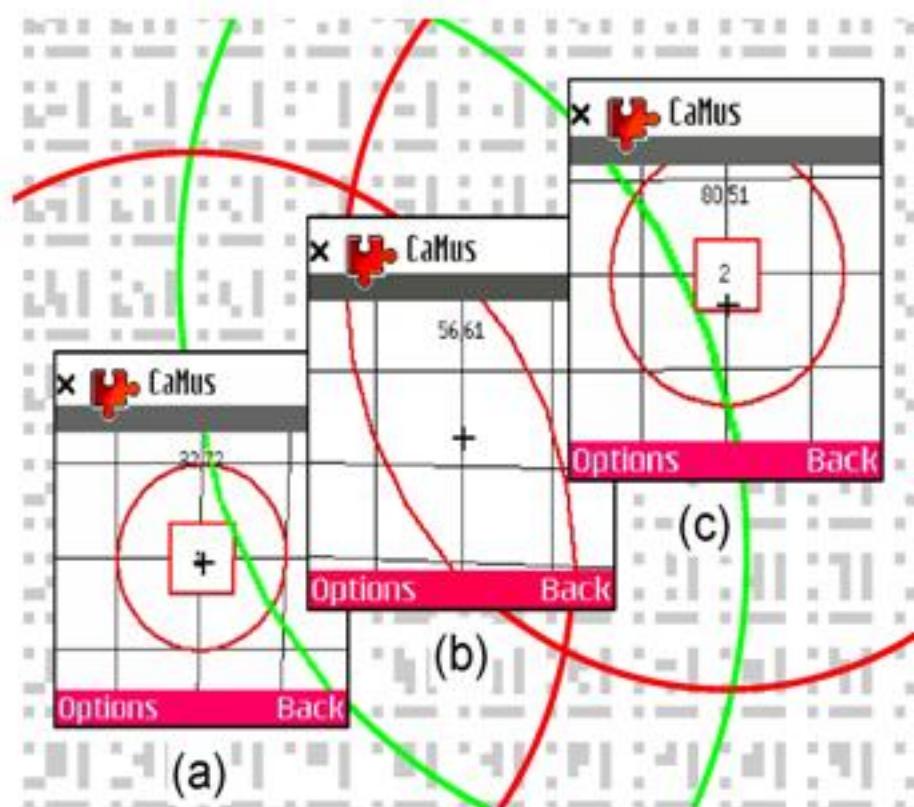
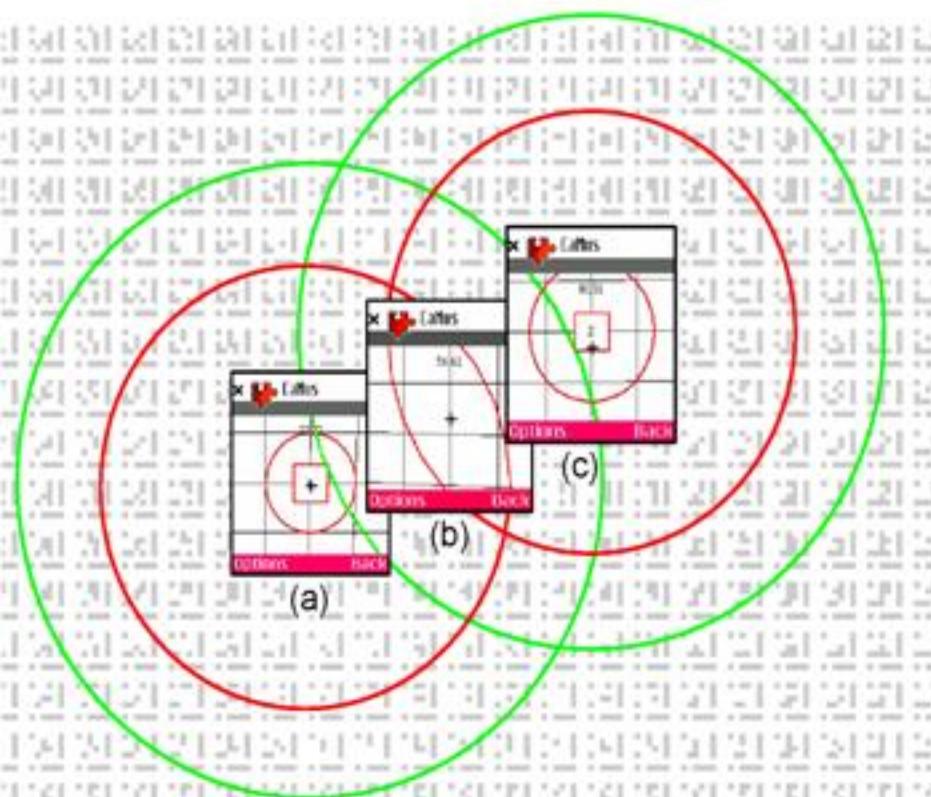
a) target "1"
locked to cursor



b) target "2"
unlocked
(default state)

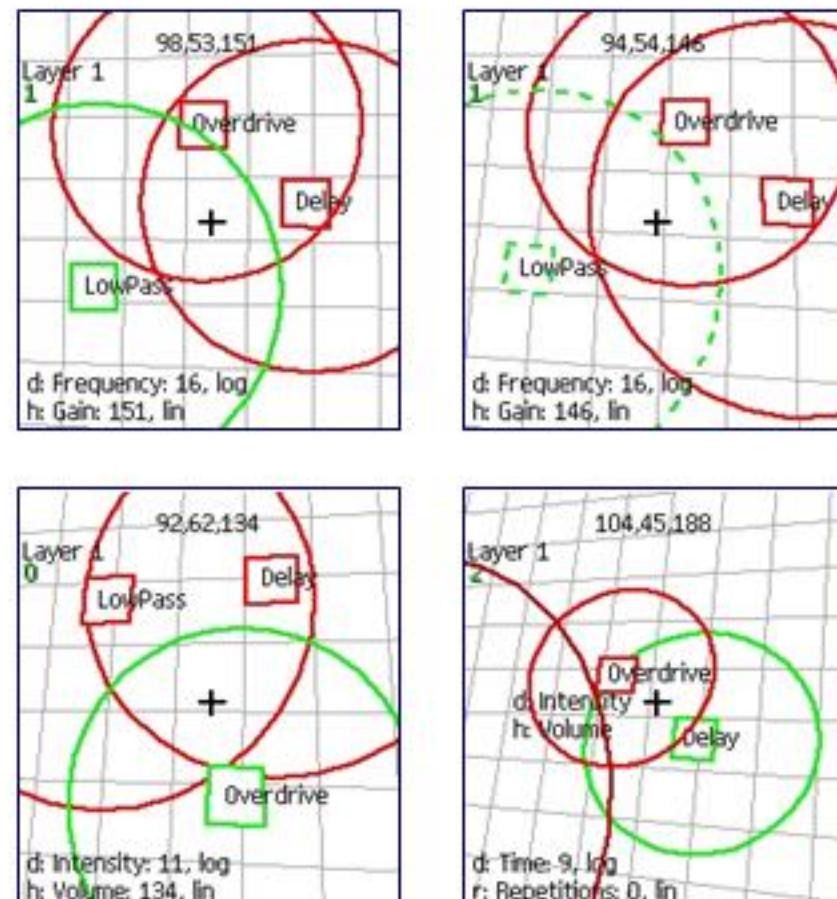
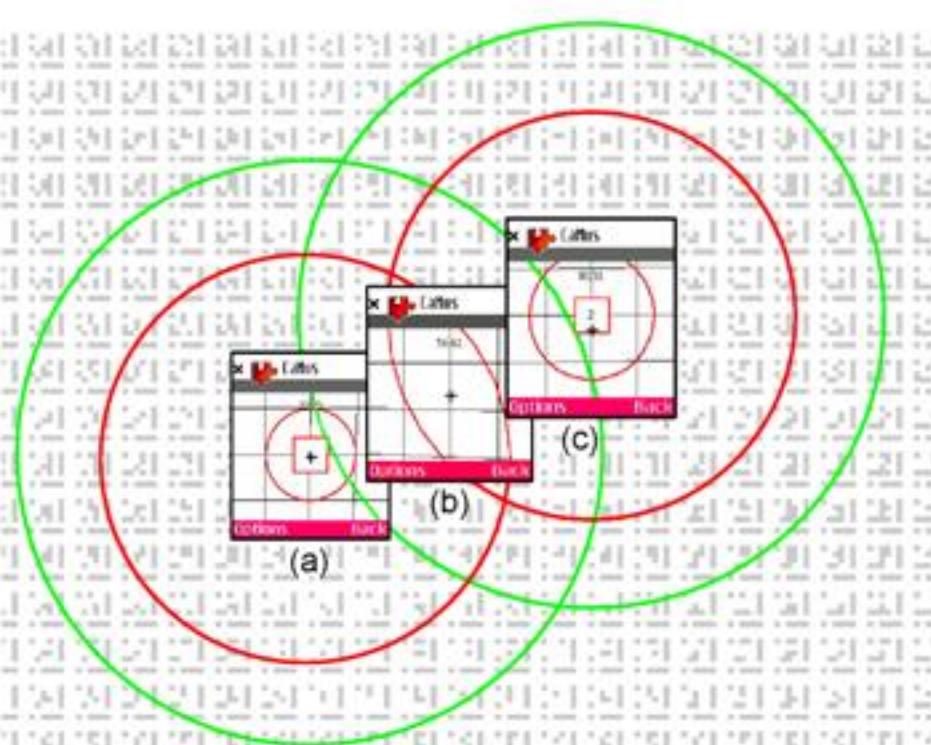
Interaction Paradigm and Visualization

- Arm range vs Visual content
- Halos visualize off-screen objects
- Halos have a fixed grid position



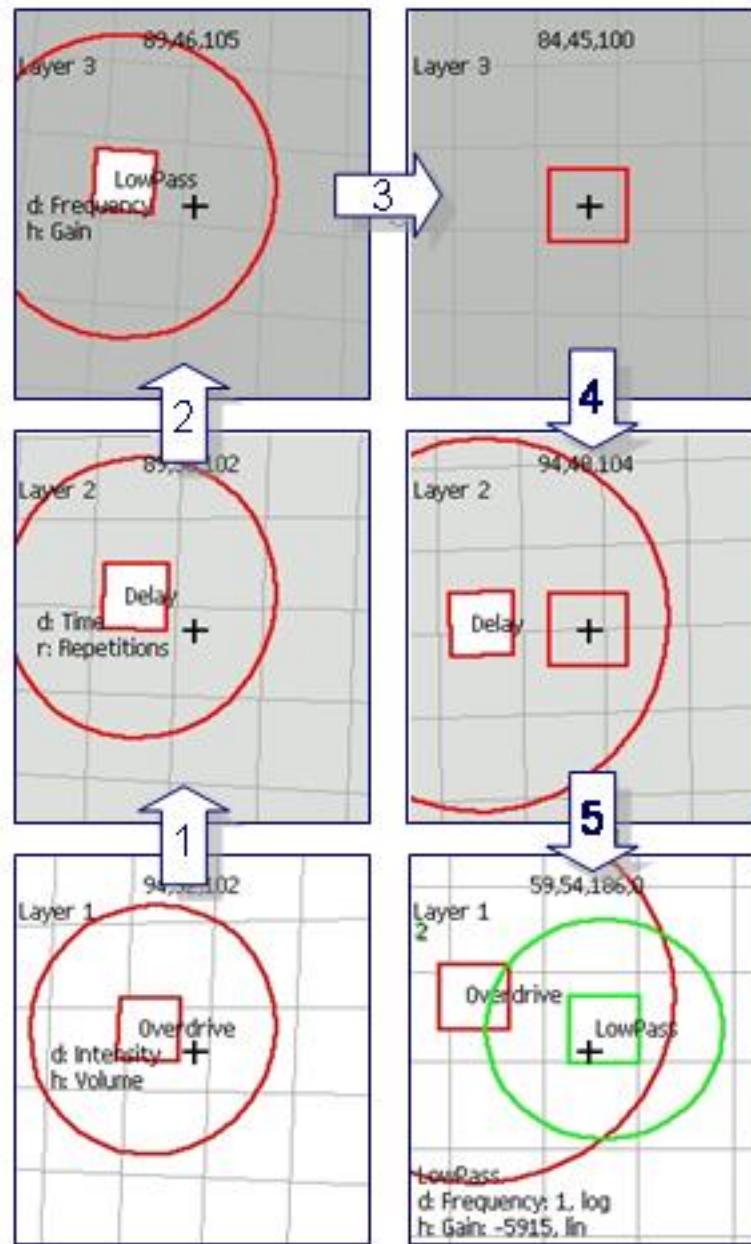
Extending the Interaction Space and “Scoring”

- Objects on grid represent sound effects or sources
- Gravity and Semantics



Multiple Layers of Interaction

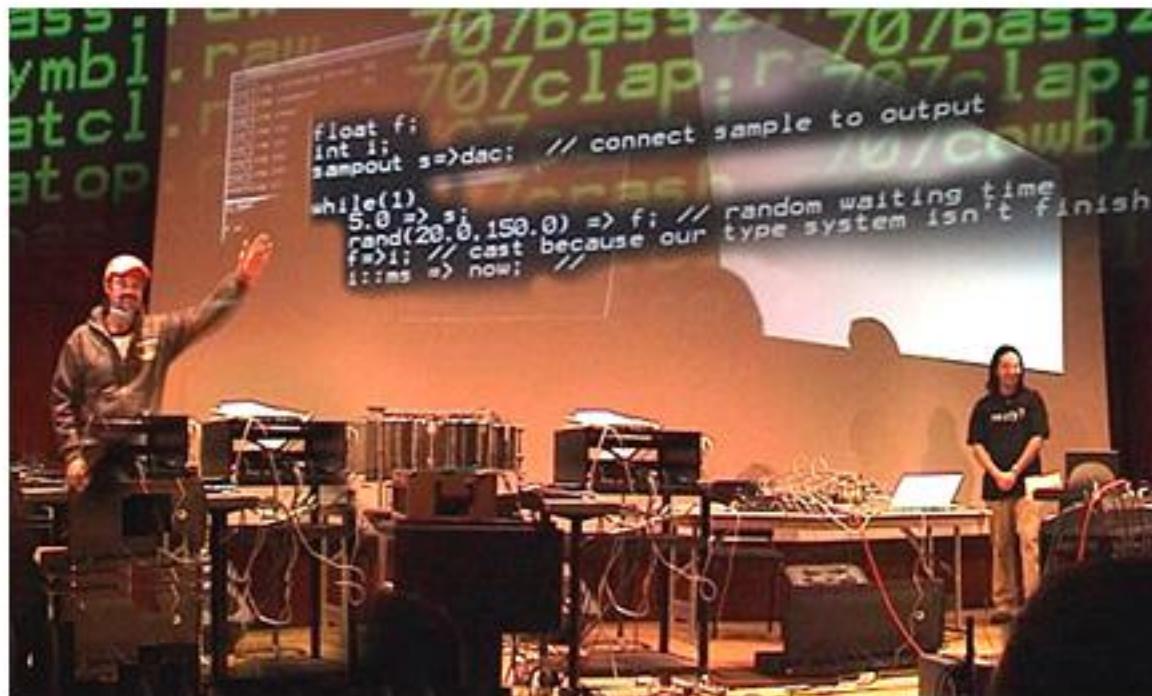
- Static vs dynamic in one score
- Switching between layers using distance gesture
- Dragging objects between layers
- Objects on other layers are unaffected by interaction on current layer



Laptops and Mobile Device.

- What's different?

- Portable/Mobile
- Untethered
- Networked?
- Commodity
- Gestural performance
- Computational power
- Quality Available Infrastructure/Software



Laptops and Mobile Device.

- What's different?

- Portable/Mobile
- Untethered
- Networked?
- Commodity
- Gestural performance
- Computational power
- Quality Available Infrastructure/Software



Conclusions.

Mobile Music Making.

- Mobile Music Making is emerging
- First technical possibilities exist
 - Key based
 - Pen/Touch screen
 - Camera tracking
 - Accelerometers
- Sophisticated sound synthesis on mobile devices exists
 - Infrastructure is still lacking, performance still an issue
- Do we need new paradigms? Visual? Gestures?
Programmability? Connectivity?

Acknowledgements.

Michael Rohs (joint on all work)
Martin Roth (first CaMus project)
Ananya Misra (ShaMus mic/voice)
Ferry Lauda (Fixed point code)

Stephen Hughes
Sile O'Modhrain
Jussi Angesleva

Atau Tanaka
Enrico Constanza
Martin Kaltenbrunner
Joe Paradiso

Henri Penttinen
Jarno Seppänen
Rob Hamilton
Ge Wang

Perry Cook
Gary Scavone

Thank you!

Georg Essl

Deutsche Telekom Laboratories, TU Berlin

georg.essl@telekom.de

Title.

Subtitle.

- Sensors

- Keys (commercial products)
- Camera (CaMus)
- Speaker (all)
- Microphone (unpublished, joint with A. Misra, Princeton University)
- Touchscreen (PDA, Geiger)

- Emerging

- Accelerometers (ShaMus)
- Magnetometer (ShaMus)
- Gyroscopes (soon)
- Capacitive (?)

Related Work – Sensor-Based

- PDA, custom sensor technology
 - Tanaka "Mobile Music Making" (2004)
- PDA, touch screen based
 - Yamauchi and Iwatake "Mobile User-Interface For Music" (2005)
 - Geiger "Using the Touch Screen as a Control Surface for Portable Computer Music Instruments"
- Camera Phone, camera based
 - Essl & Rohs "CaMus", "CaMus²" (2006, 2007)
- Parametric Synthesis
 - Geiger, "PDA" (2003)
 - Essl & Rohs, "MobileStk" (2006)



(Tanaka 2004)



(Essl, Rohs 2006)

Sensors? Integrated and Experimental.



Nokia 5500



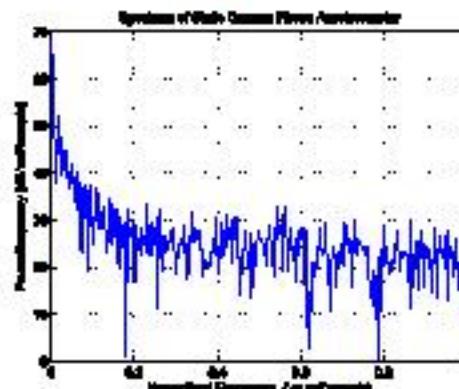
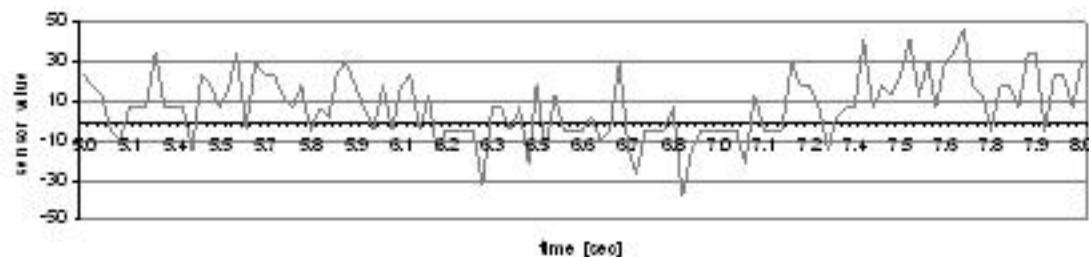
SHAKE (Hughes 2006)

Broaden sensor choices
Performance
Fidelity



Accelerometers for Striking and Shaking Gestures

- 3-axis accelerometer in Nokia 5500
 - noise over full spectrum (7% relative to static mean)
 - Use a combination of low-pass filtering and level-locking



- 3-axis accelerometer in SHAKE
 - noise (< 0.1% relative to static mean, less than 1 bit)
 - Used without signal conditioning

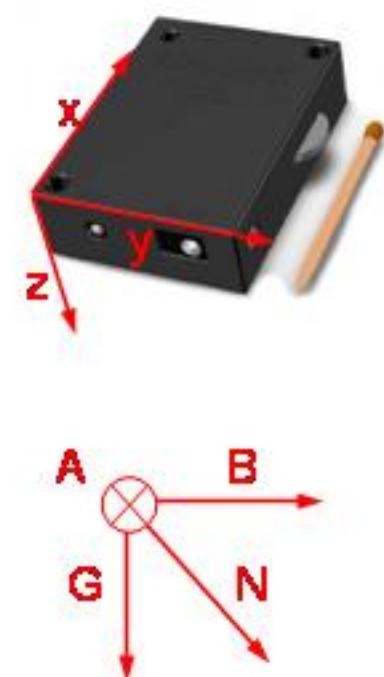
Gestures.

- Striking gesture
 - single strike event when crossing horizontal orientation
 - impact strength from local tilt gradient
- Shaking gesture
 - percussive instruments, like rattles, tambourines
 - use $\| A_n - A_{n-1} \|$ to drive physically informed particle models
- Sweep gesture (SHAKE only)
 - violin bow action
 - magnetometer for velocity
 - tilt angle for bowing pressure



Accelerometer & Magnetometer for Sweep Gesture

- SHAKE unit
 - 3-axis accelerometer and magnetometer
 - combined accelerometer and magnetometer for absolute heading and inclination
 - $A = G \times N$ and $B = G \times A$ in the horizontal plane
 - system G , A , B is orthogonal
 - heading is projection of x -axis into horizontal plane
 - inclination is $\arccos((0,0,-1) \cdot G / |G|)$
- Sweep gesture
 - violin bow action
 - magnetometer for velocity
 - tilt angle for bowing pressure



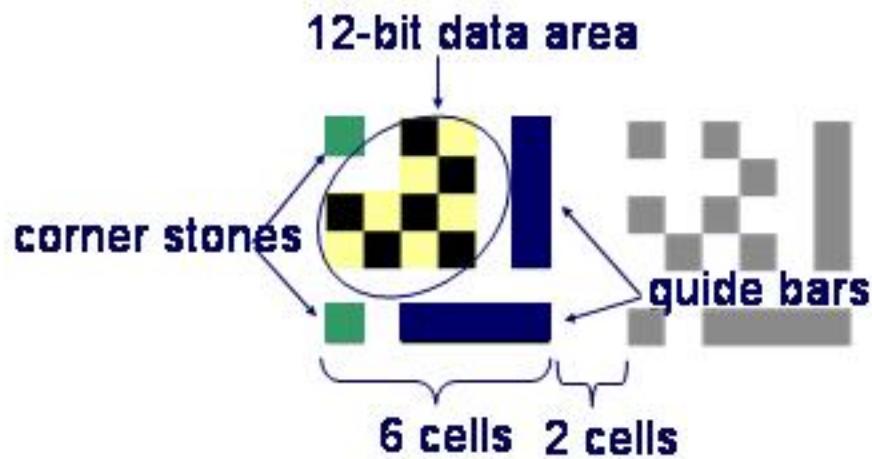
Mobile Devices with Sensors for Active Music Performance

- Contemporary mobile music paradigm
 - mobile music players / multimedia devices
 - passive, i.e. only playback
- Active mobile music paradigm
 - make the mobile device "embody" a musical instrument
 - support flexible authoring and control of mobile music
 - support non-expert users
- Sensors on commodity devices
 - sensor properties in relation to the interaction
- Design requirements for different activities
 - composition
 - interpretation
 - improvisation

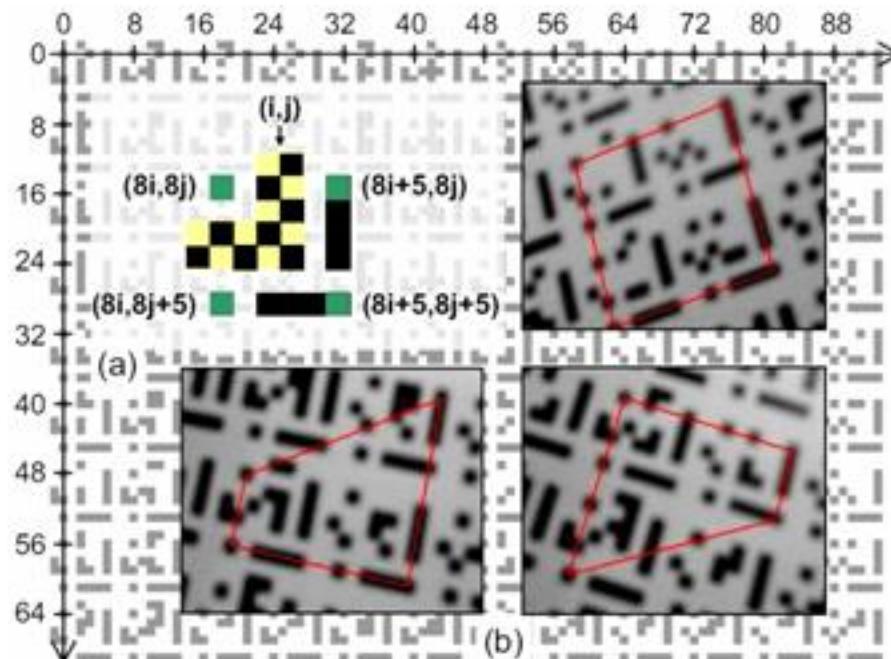


Visual Grid Tracking and Perspective Rendering

- Visual marker layout



- Markers are placed in 8x8 cell grid
- Small size ensure that always ≥ 1 marker in camera image
- Marker encodes x- and y-index within grid
- Grid defines a global coordinate system
 - used to draw graphics in grid plane



Sensing Methods for Mobile Devices

- Visual grid tracking with extended range

- absolute sensing in grid area
 - composition



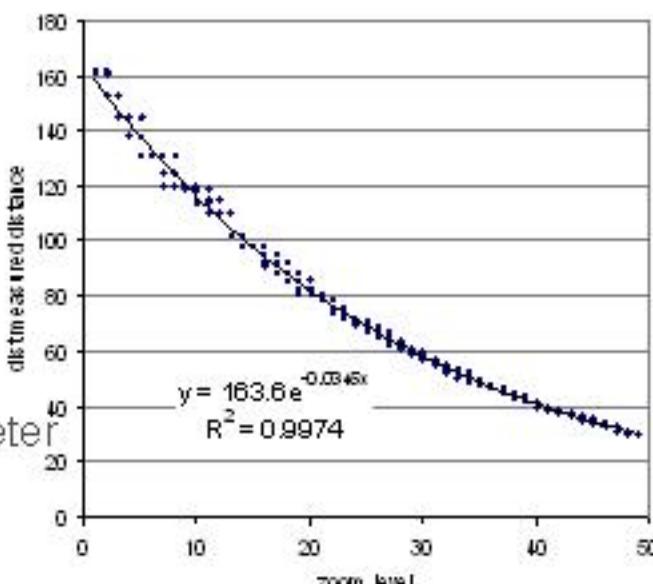
- Visual movement detection

- relative sensing on arbitrary textured background
 - composition

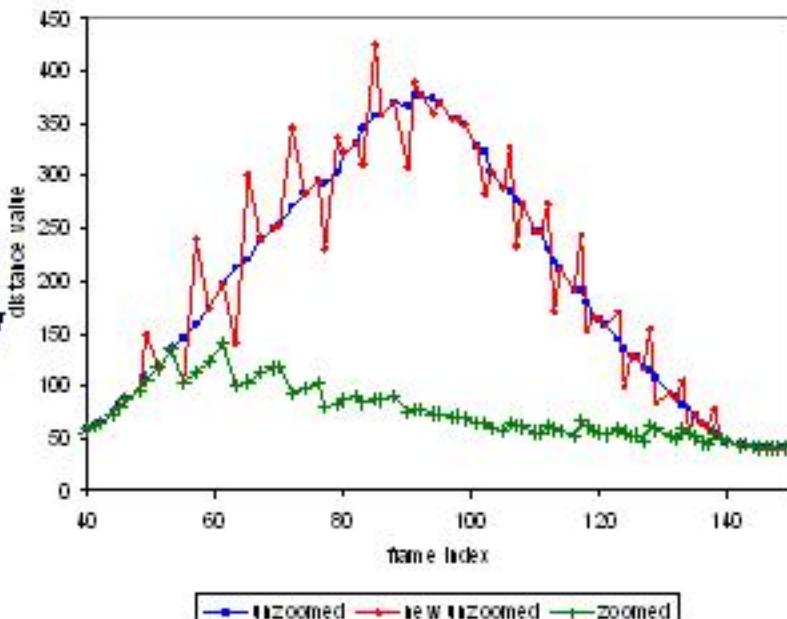


Visual Grid Tracking with Extended Range

- Dynamic digital zoom
 - continuously update digital zoom to keep optimum recognition conditions
 - $d_{zoomed}(\text{level}) = a \exp^{-b \text{ level}}$ estimated known parameter
 - $d_{zoomed}(0) = a = d_{unzoomed}$
 - $d_{unzoomed} = d_{zoomed}(\text{level}) \exp^{-b \text{ level}}$

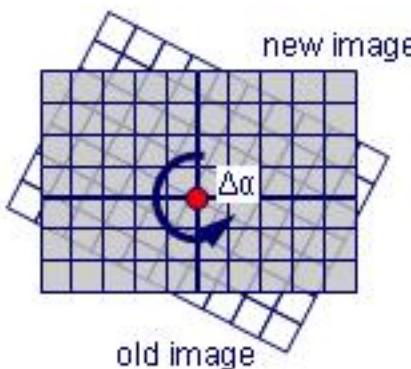
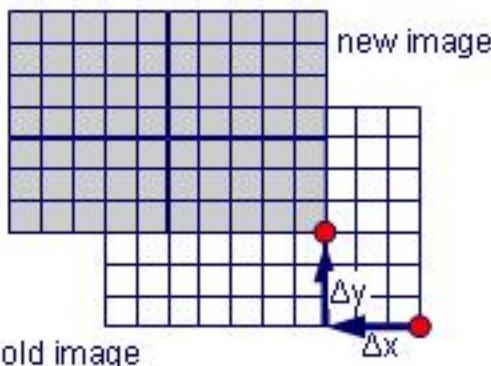


- New zoom setting not immediately valid
 - unpredictable delay of 2-5 frames
 - compute distance for old and new setting
 - choose smoothest curve



- Recognition range increases

Visual Movement Detection for Authoring



- Subdivide 176x144 pixel image in 22x18 blocks (8x8 pixels/block)
- Compute cross-correlation between successive blocks
 - frames are 67 ms apart (at 15 fps)
 - sample spacing: 4 pixels

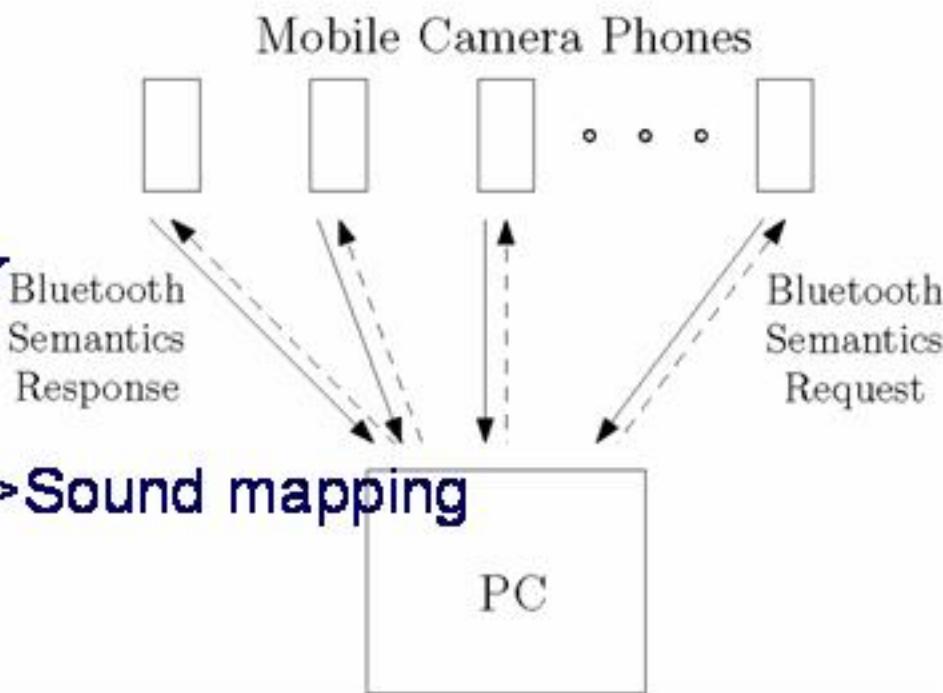
—**Typical range of linear (Δx , Δy) and rotation ($\Delta\alpha$) offsets**

$$r_t(dx, dy) = \frac{\sum_{y=0}^{h-1} \sum_{x=0}^{w-1} b_1(x, y)b_2(x + dx, y + dy)}{(w - |dx|)(h - |dy|)}$$
$$(\Delta x, \Delta y) = \operatorname{argmax}_{dx, dy \in \{-4, \dots, 4\}} r_t(dx, dy)$$

$$r_r(\alpha) = \frac{\sum_{y=0}^{h-1} \sum_{x=0}^{w-1} b_1(x, y)b_2(\text{rotate}(\alpha, x, y))}{\text{number of overlapping blocks}}$$

Local Network Collaboration.

- Current Model
 - Client-server
- Laptop/PC
 - Host Synthesis Software or MIDI engine
 - Host semantics
- Semantics and Gesture->Sound mapping are one unit



Network Protocol.

■ Symmetric Protocol

- Connection model independent
- Democratic (each participant can initiate requests)
- Ready for server-client models, client-server-client model and multiple peer model

Opcode	Description
CONNECT	Inform a host about a connect with initial information about the phone
SEND_MOVEMENT	Send movement information (x-y position, height, rotation, tilt)
SEND_PARAMETERS	Send auxiliary parameters (parameters relative to targets, x-y distance, height, rotation, tilt)
DELETE_TARGET	A target has been deleted
REQUEST_SEMATRICS	Request semantic information from other participants in the network
SEND_SEMATRICS	Send semantic information about targets (name, range and mapping of parameters)
NO_SEMATRICS	No semantic information is available for this participant

Outline

- Design requirements for different music-related activities
- Sensors and sensing methods for mobile devices
 - visual grid tracking } authoring and control
 - visual movement detection } striking and shaking gestures
 - accelerometer } sweeping gestures
 - accelerometer & magnetometer
- Design space of sensors for mobile music performance
- Comparison of sensors in the design context
- CaMus: camera phone music performance
- Conclusions

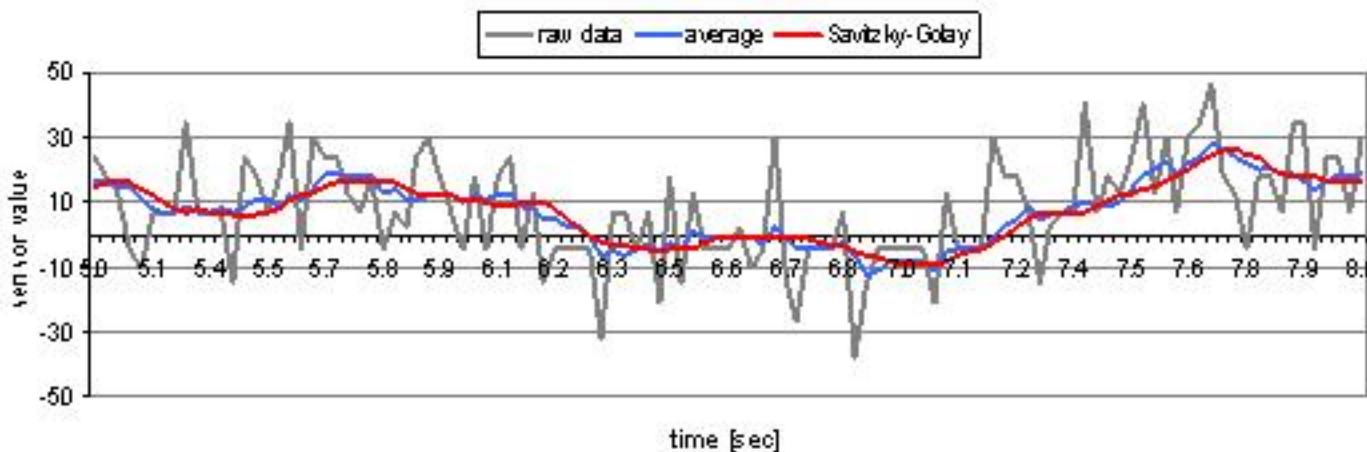
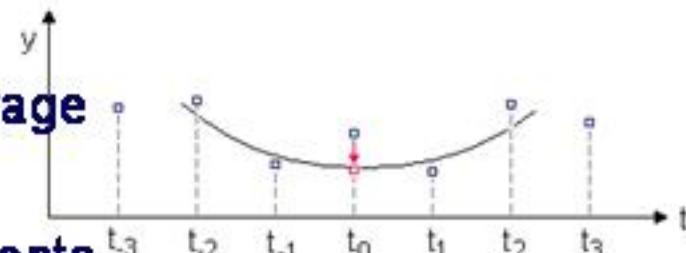
Characteristics of Sensors

Name	Type	Resolution	Sample Rates	Range	Noise	Reliability
Accelerometer	Electromechanical	High (1 mg [7])	0.5-2 kHz	±6g	Low	High
Magnetometer	Electromagnetic	High (1 mGauss [7])	1 kHz	±2 Gauss	Medium	Medium
Gyroscope	Electromechanical	High (0.1 deg/s [7])	≥ 1 kHz	±500 deg/s	Low	High
Marker/grid tracking	Optical	High [11]	15-30 Hz	150×150×30 cm	Low	High
Movement detection	Optical	Medium [11]	15-30 Hz		Medium	Low
Touch screen	Electromechanical	High		4×5 cm	Low	High
Capacitive proximity	Electrostatic	High	1 kHz [7]	0-10 mm	Low	Medium

Name	Context	Constraints	Cost	Commodity
Accelerometer	Jolt-free environments	Drift	Low	Beginning
Magnetometer	Low EMI	Requires calibration	Low	Low
Gyroscope	Any	N/A	High	Very Low
Marker tracking	Sufficient lighting	Distance, range	Medium	High
Movement detection	Sufficient lighting	Velocity, drift	Medium	High
Touch screen	Any	Screen size	Medium	High
Capacitive proximity	Low EMI	Drift	Low	Low

Accelerometer & Magnetometer Filtering

- Efficient Savitzky-Golay filters
 - retains peaks better than sliding average
 - fitting data values to a polynomial
 - convolution with fixed integer coefficients
 - reduce delay by partial evaluation



Mapping: Motion to Parameters

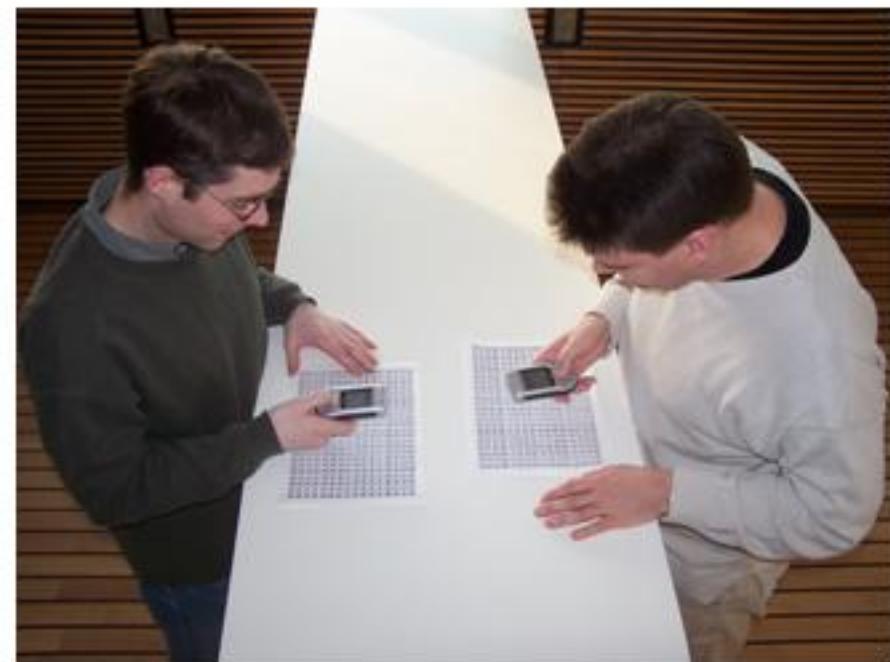
■ Mapping strategy

- General motion to most significant effect parameter
- Height mapped to effect intensity
- Rotation set to secondary effect parameter (if available)

Effect	Distance	Height	Rotation
Distortion	Distortion	Effect Weight	Not used
LP Filter	Cut-Off Freq.	Effect Weight	Not used
Balance	Left/Right	Effect Weight	Not used
Delay	Forward delay	Effect Weight	Feedback delay
Reverb	Reverb	Effect Weight	Room Size

Motivation

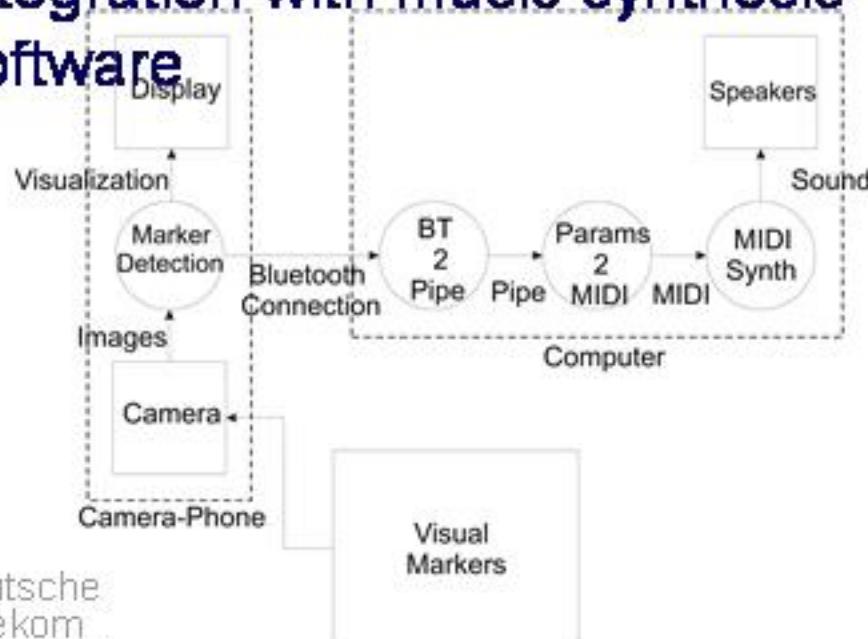
- Make the mobile device a musical instrument
- Contemporary mobile music paradigm mostly:
 - passive, i.e. only playback
 - record/playback
 - limited interactions



- Affordance on the device is generating sound inducing conventional "instrumental gestures" on the device

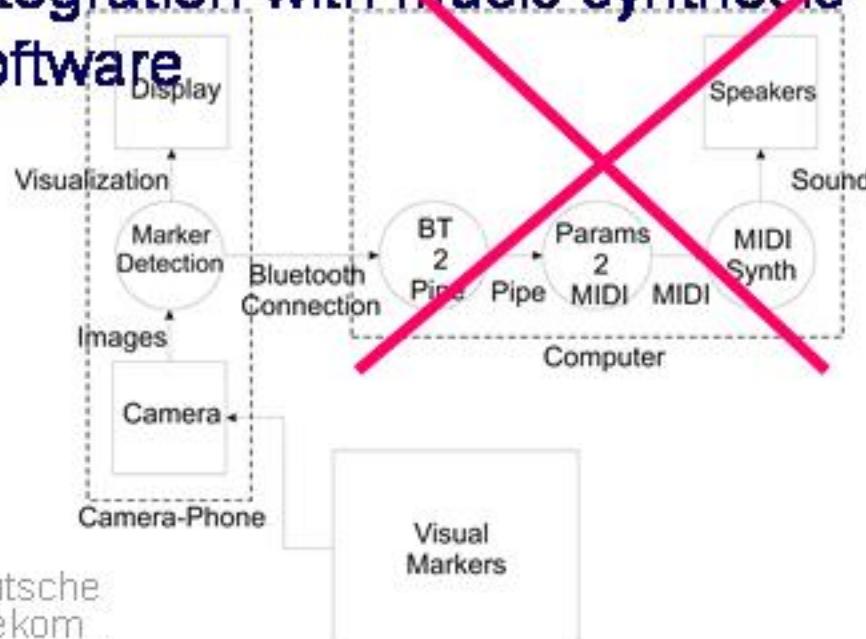
Goal (of CaMus)

- Mobile technology used for interactive real-time sound generation applications
- Visual marker grid as interaction space for mobile devices
- Bluetooth communication for integration with music synthesis software



Goal (of MobileSTK)

- Mobile technology used for interactive real-time sound generation applications
- Visual marker grid as interaction space for mobile devices
- ~~Bluetooth communication for integration with music synthesis software~~



- Put the music back on the phone

Physical and Parametric Models for Sound Synthesis

- Sound
 - Physical Models of Sounding Phenomena
- Expressive
- Cheap (Memory, CPU load)
- Efficient
- Physically Relevant Parameters
- Real-time, Interactive

STK. Portable Library for Parametric Sound Synthesis

- Developed by Perry Cook (Stanford, Princeton) and later Gary Scavone (Stanford, McGill) 1990+
- C++, portable
- Focus on contemporary parametric methods (often immediate research)
- Focus on real-time parametric synthesis and physical models.

Symbian OS Oddities and STK.

- Lacks global static memory.
 - `stk::sampleRate()`: Revert to `#define SRATE`
 - Error messaging: Scrapped
 - Shaker index: made local to class
- Lack of C++ Std template libraries
- Symbian OS Audio
 - `CMdaAudioOutputStream`
 - `MMdaAudioOutputStreamCallback`: Called when relevant events happen, like a stream got started, a buffer is emptied, a stream is closed.
 - No locking or copying of stream buffers: Double buffering
 - No or very limited MIDI support. No RTMidi for Symbian.

Ported Features of STK.

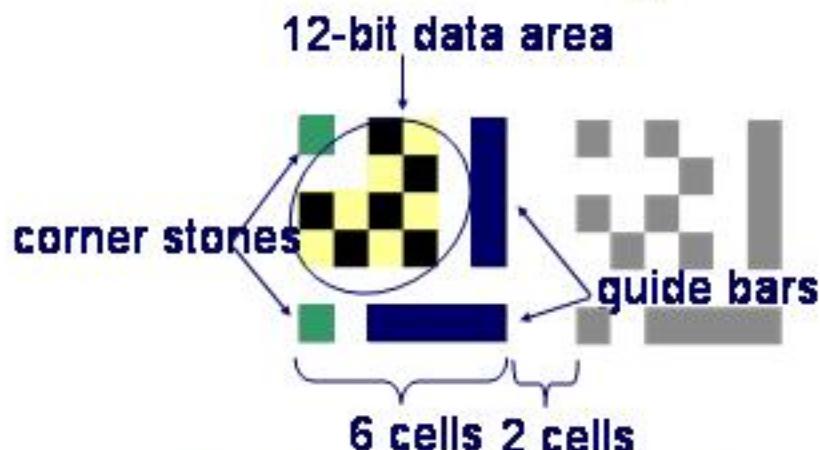
- **Support without modification:**
 - All signal processing and instrument classes are fully ported.
 - File reading and writing support included.
- **Modifications to accommodate lack of std libraries**
 - string -> const char***
 - Mimick implementations of all concrete versions of vector templates (symbmath.cpp/h)**
 - FloatVector
 - IntVector
 - ADSRVector
 - WaveLoopVector
 - FileWvInVector
 - GrainVector

Performance.

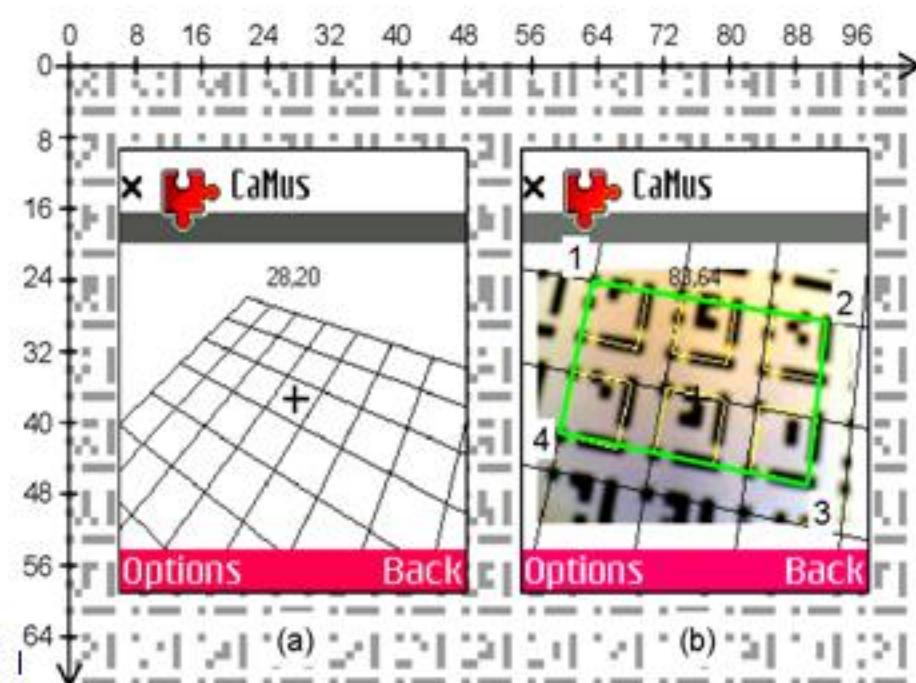
- Nokia 6630: 10+64MB, 32-bit RISC CPU based on ARM-9 series, 220 Mhz
- Plucked and SineWave at increasing numbers
- 8000 Hz, Buffer size 1024
 - 5 plugged strings, 9 sine waves
- Same for 2048 and 8192 -> computational power is bottleneck
- 16000Hz, Buffer size 1024
 - 2 plugged strings, 4 sine waves

Visual Grid Tracking and Perspective Rendering

- Visual marker layout

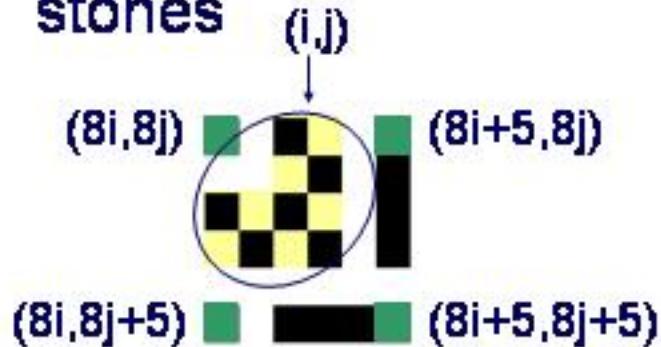


- Markers are placed in 8×8
- Small size ensure that always ≥ 1 marker in camera image
- Marker encodes x- and y-index within grid
- Grid defines a global coordinate system
 - used to draw graphics in grid plane

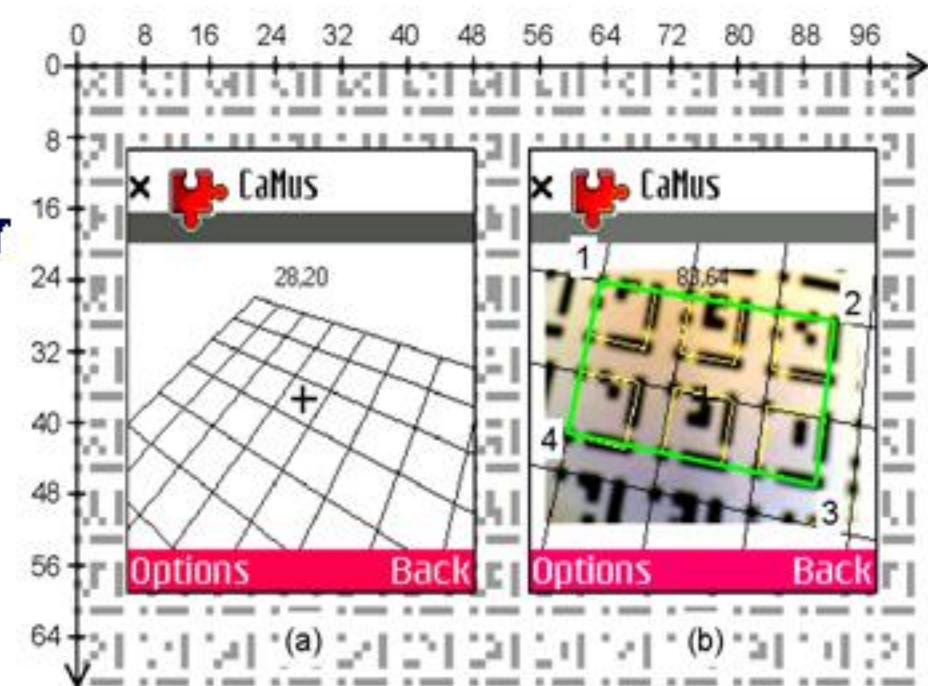


Visual Grid Tracking and Perspective Rendering

- Marker encodes index position (i,j) within grid
- Grid coordinates of corner stones



- Four corresponding point pairs required for perspective mapping
- Use multiple markers if possible,
for stability of mapping



Element (marker corner)	Image coordinates (from marker recognition)	Grid coordinates
Left top	(x_1, y_1)	$(8i_1, 8j_1)$
Right top	(x_2, y_2)	$(8i_2 + 5, 8j_2)$
Right bottom	(x_3, y_3)	$(8i_3 + 5, 8j_3 + 5)$
Left bottom	(x_4, y_4)	$(8i_4, 8j_4 + 5)$

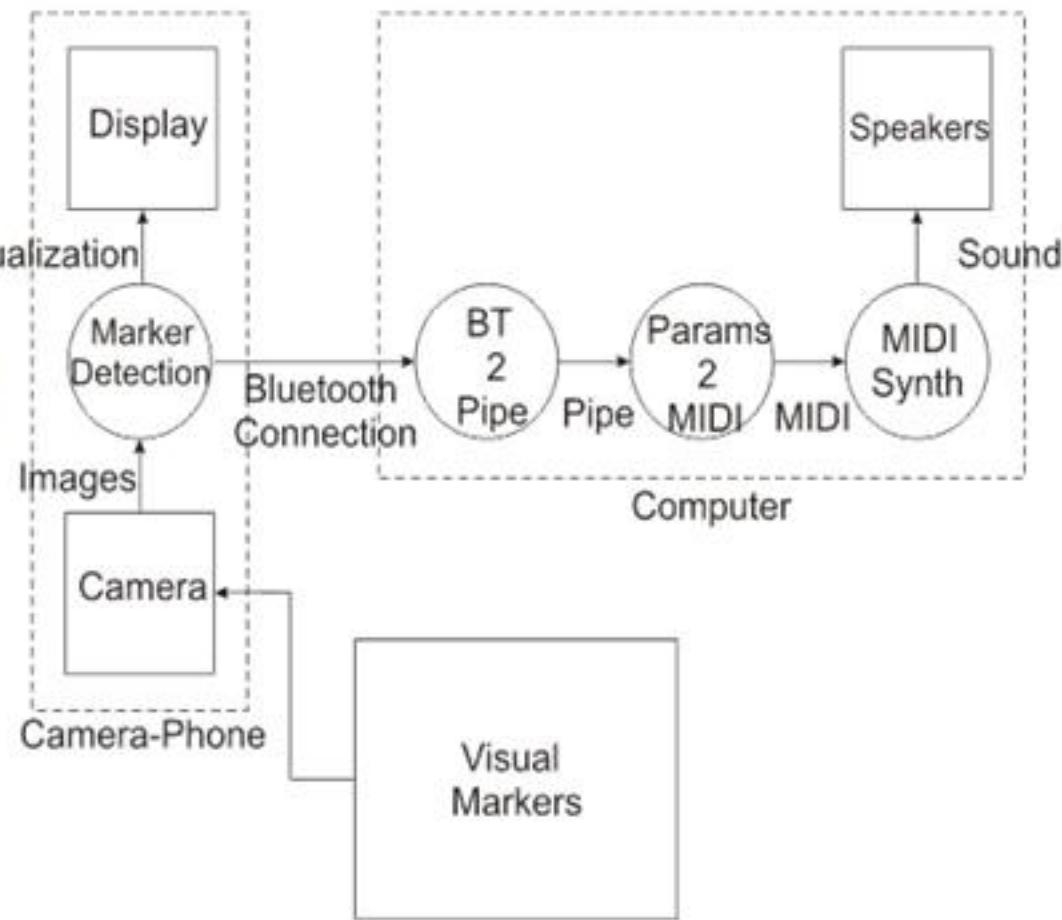
Mapping Motion to Music

- Data sent over Bluetooth

- device rotation relative to grid
 - device distance to grid
 - grid coordinates of cursor (screen center)
 - grid coordinates of all targets

- Update rate

- 5-10 updates per second
 - on current phone hardware (Nokia 6630, Symbian, 10 MB



Mapping: Sound Synthesis

- MIDI to Effect mapping
- On-the-fly modification of sequence music
- Multiple parameter interaction in parallel



Mapping: Motion to Parameters

- **Mapping strategy:**

- General motion to most significant effect parameter
- Height mapped to effect intensity
- Rotation set to secondary effect parameter (if available)

Effect	Distance	Height	Rotation
Distortion	Distortion	Effect Weight	Not used
LP Filter	Cut-Off Freq.	Effect Weight	Not used
Balance	Left/Right	Effect Weight	Not used
Delay	Forward delay	Effect Weight	Feedback delay
Reverb	Reverb	Effect Weight	Room Size