ANCHORING IN UBIQUITOUS MUSICAL ACTIVITIES

Damián Keller*  
Universidade Federal do Acre  
Núcleo Amazônico de Pesquisa Musical - NAP

Daniel Luís Barreiro  
Universidade Federal de Uberlândia  
Núcleo de Música e Tecnologia

Marcelo Queiroz*  
Universidade de São Paulo  
Grupo de Computação Musical

Marcelo Soares Pimenta*  
Universidade Federal do Rio Grande do Sul  
Laboratório de Computação Musical

ABSTRACT
Prevalent music techniques are not well suited to music making in ubiquitous contexts. How are musicians and system designers supposed to find their way through the myriad of locations, devices, and materials involved in ubiquitous music interventions? This study suggests that we may get some insights into the mechanics of ubiquitous music making by looking at the intersection between tools, concepts and activities. We discuss affordances as dynamic properties emergent from agent-object (natural affordances) and agent-agent interactions (social affordances). And introduce anchoring as a conceptual tool to explain how affordances are shaped. Then we apply anchoring in the development of a sound-based interaction metaphor - time tagging. We report the implementation of a prototype – mixDroid – and a usability test involving the realization of a musical work.

INTRODUCTION
This paper introduces the concept of anchoring within the context of ubiquitous music systems design. Anchoring has been utilized in several disciplines including robotics [3], psychology [7, 15] and music [11, 12]. We discuss its applicability in ubiquitous computing and human-computer interaction.

The first section of the paper defines the objective of our study within the context of ubiquitous music research. Ubiquitous music [10, 21] has been proposed as an area of research that focuses on the intersection of pervasive computing (ubicomp [25]) and music (figure 1). Previous work on installation art, mobile music, cooperative composition and ecocomposition provides support for this emergent field. Affordance is one of the concepts that have been applied both in interaction design and in ecocomposition. We point to the limitations of the use of affordances in current HCI research and suggest a mechanism for affordance formation.

Figure 1. Ubiquitous music [10, 21].

In order to test the applicability of anchoring in the context of a ubiquitous musical activity, we focused on one of the most basic uses of technology for sonic work: mixing. Mixing was chosen for three reasons. First, it encompasses a set of actions that can be isolated, quantified and analyzed as a coherent whole without disrupting the broader context of music-making activities. Second, the actions executed in mixing can be supported by tools developed for consumer-level devices. Third, this activity affords a musically meaningful result: the sonic mix.

* Ubiquitous Music Group (Ubimus), ubiquitousmusic@googlegroups.com.
As a pilot study, we approached the design of a ubiquitous interface for mixing. The study involved the implementation of a prototype that supported a stripped-down definition of the activity at hand. By mapping a series of snapshots of a real-world sonic experience onto the constrained space of the mixing procedure, we obtained a metaphor that captures the minimum variables necessary for the study of mixing: time tagging. The variables controlled in this interaction metaphor are time tags and sound samples. This simplified definition fulfilled two requirements. On one side, it was simple enough to provide the basis for an easily scalable interaction technique. On the other, it preserved the musical complexity of the process.

The validation process was as close to actual usage as a rapid prototype would allow. The objective of the validation was to test whether the interaction technique would support a complete cycle of activity: from the initial state – a collection of unordered sound samples, to the final state – a spacetime organized set of sounds. The subject chosen for the study had full musical training and over twenty years of experience with technology. This choice was based on the need for a stringent assessment of the quality of the musical result. If the interaction technique did not support a high-quality musical product, this avenue of research would not justify further exploration and should be aborted at an early stage.

**Ubiquitous Music: An Uncharted Territory**

As composers and system developers, ubiquitous musical activities present us a completely new and uncharted territory. Previous musical practices provided the safe refuge of instruments as the physical support for all sound-producing actions. These actions could be encoded as a series of discrete symbols - a score - which would guide the performers through a finite set of possible interactions with their instruments. Performances would occur within a space especially designed for musical activities - the concert hall - guaranteeing acoustic characteristics compatible with instrumental sound source power and projection. Furthermore, a crisp separation between performers and public, following an established ritualized set of actions - play / listen, bow / applaud - reinforced by the physical separation between stage and audience seats, allowed for strictly predefined roles in music making: musicians play, spectators just listen.

Most of this social paraphernalia breaks down in the context of ubiquitous music practices. Instead of clearly identifiable musical instruments, we have a multiplicity of interconnected devices which may serve both as sensors and transducers of sonic, visual and haptic data. In ubiquitous musical systems sonic data is transformed between the analogue and the digital domains but it is not necessarily restricted to a preestablished set of discrete symbolic representations. More important, haptic actions do not map directly onto sonic results. So what you touch is not what you play and what you see is not what you hear. On the same token, sound source locations and effectors do not necessarily follow one-to-one mappings. So where the action takes place is not always where the sound will be heard.

This already complex picture is further complicated by the fact that all agents participating in a ubiquitous musical activity may be able to influence the sonic results. Given that spatial separation is afforded through remote connectivity, musicians and public need not be in the same place at the same time and they do not even need to be hierarchically or functionally separate entities. In fact, some ubiquitous systems support music making as a social activity where roles are not predefined, and experts coexist side-by-side with newbies [18]. In this context, playing, listening, composing are all intertwined with a larger set of social interactions. Within this broader community process, music becomes just a sonic byproduct of the main focus of the activity: the extramusical interactions.

Given that most preestablished music-related social rituals are not well suited to the ubiquitous music milieu, how are musicians and system designers supposed to find their way within the fleeting myriad of locations, devices, and materials involved in ubiquitous interventions? By looking at the intersection between tools, concepts and activities we may be able to get some insights into the mechanics of ubiquitous music making. Let’s start by discussing the concept of affordance and how it relates to human-computer interaction.

**Natural Affordances or Social Affordances?**

One of the first applications of Ecological Psychology [6] to HCI design was done by Vicente and Rasmussen (1992) while working at the Risø National Labs in Roskilde, Denmark. Given the direct relationship between human actions and physical objects, Vicente and Rasmussen proposed three principles for interface design: 1. Control through direct manipulation, involving the use of a spacetime representation to support direct interaction; 2. Consistent mapping between the work domain and the actions supported by the interface; 3. Support for an externalized mental model of the work space by an abstraction hierarchy. The first two strategies have been adopted by several researchers in the field of interaction and are consistent with Ecological Psychology and Activity Theory [16]. The third item points to a limitation common to the Risø and the standard usage of ecological concepts in the context of HCI: the adoption of a representation mechanism that would mediate internal and external processes [2].
Are social affordances just natural affordances applied to multiple agents? This seems to be the standard view in HCI research [5, 20]. The source of the confusion is the common mechanisms for agent-object interaction and agent-agent interaction.

Natural affordances arise from an iterative process where the agent acts on the environment and both environment and agent are modified through mutual adaptation. When other agents in the same ecological niche pick up information resulting from the pragmatic actions exerted by an agent, social interactions are established.

Aside from actions on objects, natural affordances are also determined through cognitive processes that take place within the personal environment and which are directly linked to action-perception. The ongoing experiences with objects in the environment provide spatiotemporal referents that shape cognitive operations. External objects are used to situate abstract concepts. By manipulating spatial and temporal external points of reference – or material anchors – cognitive processes can make direct use of environmental information. This process is called projection [15].

Each agent has its own set of natural affordances and this set determines the agent's personal environment at a given spacetime frame. Social interactions result from the externalization of natural affordances but there is no guarantee that all agents will share the same personal environment. In fact, genetic diversity and different individual experiences may foster divergent adaptation mechanisms to the same local conditions.

Contrastingly, social affordances are directly dependent upon actions on the social environment, thus they are constructed exclusively through pragmatic actions. In other words, cognitive processes can only become social interaction when they embody externalized actions. This means that for a given individual, natural affordances and social affordances are created through processes which share a common environmental niche but do not necessarily share the same dynamics.

The error of the standard HCI view is to equate the personal sense / environment with the social environment. Social interactions are just a subset of the interactions of an agent with her surroundings. As a complement, the personal sense encompasses agent-object interactions and affordances determined by individual cognitive-propraoceptive processes which may not be directly accessible to community-based interactions. Thus, designing for social affordances only guarantees the ecological validity of the interaction metaphor within a specific community or social niche. And on a similar vein, metaphors based exclusively on natural affordances do not necessarily take the socially mediated mechanisms into account.

Contrary to the standard HCI interpretation of affordances [20], for Riso researchers affordances are specific to human work activity and involve socio-historical, cultural and organizational dimensions, goals and constraints. According to their approach, affordances evolve dynamically, are embedded into socio-cultural contexts and are not essentially nor causally given. Rasmussen and coworkers (1994) defined affordances as cues for action relevance. But due to the necessity to contextualize the use of affordances, this initial definition was modified to "cues for action relevance, displayed in the context of a virtual ecology of work" [1].

Gibson’s (1979) view of the environment focuses on the coevolution of agents and objects and departs from the mechanistic and representationalist explanations of psychological phenomena. Within the ecological perspective, information systems are not abstract representations of natural environments, they constitute particular forms of materiality whose creation, through manufacturing and design, involve high-order invariants. These invariants are externalized through affordances and encompass both structural and functional coadaptations. As natural affordances evolve through interactions with the environment, information systems need to adjust to new ecological niches. Therefore, the invariants of the ecology of distributed activity are not essentially given, but coevolve through multilevel couplings between agents and objects within specific activity domains.

**Anchoring**

“It is a mistake to assume that thinking is, in general, a symbol manipulation process” (Hutchins 2005).

Affordances are not properties of the environment or properties of the actors. They are relational properties that arise while activities are being carried out. Activities involve cognitive and proprioceptive processes that engage both external objects and conceptual operations. By understanding affordances as dynamic properties emergent from agent-object (natural affordances) and agent-agent interactions (social affordances), a previously overlooked area of HCI becomes the focus of the research agenda: how affordances are shaped. Anchoring is one of the key mechanisms for cognition and proprioception integration. We believe that it also plays an important role in affordance formation.

For Hutchins (2005), mental processes require stable representations of constraints. He suggests two ways to achieve stability in conceptual models. Cultural models
achieve stability via a combination of intrapersonal and interpersonal processes, i.e. through social interaction. Conceptual models that are well established within a community use social interaction as a grounding strategy. The other mechanism for stable conceptual systems is the association of conceptual structure with material structure. This is called conceptual blending or anchoring.

Regarding conceptual relationships, material anchors are established when the abstract elements are mapped onto a material pattern in such a way that the perceived relationships among the material elements are taken as proxies. In this case, the material pattern is providing the handles for the anchoring process. Thus, Hutchins defines the ‘material anchor’ as the input space from which material structure is projected into a conceptual blend. Material anchors range from the minimum case of individuation - an element becomes distinguishable from all others - to complex systems of relationships among concepts.

Conceptual models must maintain a system of constraints while being subjected to cognitive or physical manipulation. The stability of some conceptual models is ensured by their simplicity. For example, one-to-one mappings between digital parameters and physical points of access provide the simplest virtual-physical object relationship. In other words, material objects can be used to fix referents that are manipulated through conceptual operations taking place in the cognitive realm.

Aside from conceptual blends, the stability of complex conceptual models is provided by conventional (culturally shared) well-learned mental constructs. Hutchins (2005) suggests that a conceptual model with these properties is equivalent to a cultural model. Whether the concept is purely mental or embodied in physical form, the principal cognitive effect of using conceptual models is achieved by establishing the constraints of the activity within the structure of the model. Thus, anchoring serves as a mechanism for linking constraints of the external structure of the environment to constraints on cognitive operations.

Projection is a form of anchoring that has been proposed as a way to reduce cognitive load in complex activities. Kirsh (2009) states that projection is similar to perception, but less tied to what there is in the environment. Projection, unlike pure imagery, requires external structure to anchor conceptual elements. The difference between imagining and projecting is that while imagination does not use physical objects as direct referents, projection demands the presence of objects to help the cognitive manipulation of conceptual elements. This view is line with Hutchins's suggestion that both memory and processing loads can be reduced if the constraints of the activity can be built into the physical structure of a material device. The problem faced by system designers is thus reduced to finding a consistent mapping between the abstract concepts and the physical elements of the interface. In other words, appropriate metaphors for projection should support direct couplings between physical and conceptual operations.

From Sampling to Mixing

Sounds occur in the context of complex activities that usually involve interactions among multiple agents and objects. The sonic byproduct of these interactions are the sonic events that compose the sound field. By focusing on isolated activities, sampling provides a temporally and spatially delimited snapshot of these ongoing interactions. As long as the sampling process remains consistent with the temporal rate of the interactions, sampled sounds will retain the time-based perceptual cues of the sources. Thus, microtemporal cues require sampling windows at a mesotemporal level. Mesotime demands samples of full-length events. Macro sampling - activity-related time - needs to be done through a full activity cycle.

Because sampling separates the sources from their original locations, it eliminates ecologically valid spatial cues. Locations of sound sources are determined by a combination of amplitude and microtemporal cues. Both interaural level differences and interaural time differences are dependent on azimuth and elevation of the source as measured from the position of the listener. Distance cues are determined by the relative levels of direct and reflected sound. The process of isolating direct sound from sound sources – through sampling - reduces all spatially related variables to a single dimension: amplitude.

Whether the sample pool - resulting from the sampling process - is composed of grains, events or snapshots of full activity cycles, the temporal reorganization of the samples implies a compositional choice. Thus, mixing necessarily entails control both of the temporal cues and the sound-level cues. Space cues and temporal cues are usually handled as independent variables in most compositional approaches (but see [8] and related techniques). Therefore, as long as microtime is kept out of the picture, mixing tools may provide independent compositional control of amplitude levels and temporal distribution of the samples.

The separation of space-based and time-based variables simplifies the system design process, reduces the learning cost of the interface and expands the opportunities for repurposing of control hardware. By focusing on the temporal control of sounds, we can define mixing as a subclass of a general compositional problem within the field of interaction. Because the methodological focus is aligned with the procedures adopted in sampling, this separation does not have a negative impact on the ecological validity of the process.
Definition of Mixing:

Mixing constitutes an activity in which an agent or agents change the state of a set of sampled sounds from a random distribution to a temporally organized state. The original state can be described as a sample pool. The final state constitutes the mix.

Time Tagging

This is where time tagging enters the scene. As an interaction metaphor, time tagging defines a process by which a set of unordered virtual elements or processes is layered onto an abstract one-dimensional structure—a tagged timeline. Time tagging involves haptic actions, perceptual cues, categorical memory, and cognitive anchors.

MixDroid 1.0 is a testbed for the application of the time-tagging metaphor to mixing. In mixDroid, time tagging proceeds by synchronous sequential access of each of the sampled sounds stored in memory. Sound data is sent to a mixing buffer which constitutes the tagged timeline. While the user accesses the samples, the contents of the buffer are being played. This mechanism provides sonic cues that serve to shape the temporal anchoring process. The user listens to the sounds on the timeline and the samples being played provide cues for the new sounds that will enter the mix. Finally, the new sound entering the timeline is tagged when the user presses a key (figure 2).

![Figure 2. Time tagging: a sonic metaphor for mixing.](image)

In time tagging, proprioceptive actions are done through points of access, i.e., discrete physical elements of the interface that can be mapped onto the virtual elements or processes to be tagged. In the case of mixDroid, we chose the mobile phone’s numeric keypad as our access infrastructure. The keypad is found on most mobile and fixed devices, therefore this metaphor can easily be translated to a multiplicity of low-cost gadgets. Mobile phones feature up to 12 keys. Since Miller’s (1956) first findings, several studies have corroborated that humans can easily hold approximately seven elements in working memory. Thus, by mapping one numeric key per sound sample, cognitive short-term memory constraints are not exceeded.

TIME-TAGGING STUDY

A difficulty faced by the designers of musical tools is the slowness of the validation cycle. Because complete integrated systems are hard to design and test, tools usually deal with isolated aspects of musical activity. Musicians’ usage of the tools may not correspond to the intended design and integration of multiple elements may give rise to unforeseen problems. As a partial solution to these hurdles, we have suggested the inclusion of music making within the development cycle [4]. This integration of music making and software development is based on a broad approach to usability. Fine-grained technical decisions are done after the usability requirements of the system have been well established through actual usage. So rapid deployment is prioritized over testing on a wide user base.

The subject chosen for the study (the first author) had full musical training and over twenty years of experience with music technology. This choice was based on the need to integrate actual expert usage within the development cycle. Because compositional decisions are hard to assess and the quality of the musical product may involve subjective evaluations, integrating composing and validation seems to be the most straightforward way to avoid these difficulties.

In order to test the usability of the time-tagging metaphor, we developed a time-tagging prototype for the Android operating system: mixDroid. MixDroid was developed in Java SE using the Android Development Toolkit. Given the limited support for sonic work on the Android OS, we had to implement tools for sound file handling and real-time streaming [21]. MixDroid provided an initial testbed for these tools. An example of use is available at http://www.youtube.com/watch?v=Mbpfaq1dcY0. Figure 3 shows a snapshot of the mixDroid 1.0 time-tagging interface.
mixDroid 1.0: an implementation of anchoring in the context of ubiquitous musical activities.

The mixing procedure consisted of several mixing sessions. MixDroid was used in the emulation mode on a laptop PC with a dual-core two-gigahertz CPU and three gigabytes of RAM. All samples were stereo WAVE format, 16 bits, 44.100 hertz. Several dozens of sound samples were used, with durations ranging from a few centiseconds to approximately two minutes. The mixDroid interface was activated through pointing and clicking with an optical mouse.

Results of the Evaluation

The validation process of the interaction metaphor involved the use of the mixDroid prototype for the creation of a complete musical work. The main objective of the validation was to test whether the interaction technique would be suitable for a complete mixing cycle: from the initial state – defined as a collection of unordered sound samples, to the final state – a time-based organized set of sounds. Although sound quality and accurate synchronization were not key aspects in this first cycle of the validation, we ensured support for standard PCM formats and sample-based synchronization.

The subject composed a seven-minute stereo work called Green Canopy: On the Road. The piece makes use of sampled sounds collected at a location in the Amazon Rainforest - the Yanayacu River, in Peru. This sonic palette was extended by means of ecological modeling [13] and asynchronous granular synthesis [23]. Resonant filtering was used to modify the spectral qualities of the sounds, keeping the macrotemporal cues of the sources untouched.

Frog vocalizations are a characteristic feature of the night soundscape in Yanayacu. Following standard procedures in ecocomposition [12], during the editing process special care was taken not to disrupt the temporal relationships present in the original recordings. Thus, the temporal structure of the piece could be based on the temporal distribution of the frog calls. Sound events were mixed following the timings of the calls.

Limitations of the Study

The study was not designed for novice users. Nevertheless, the direct mapping between points of access and sound samples may prove to be intuitive enough for users with little experience with technology. Although up to eight sound samples were employed at each of the mixing sessions, it is not clear how many elements should be available for untrained users. Another aspect that needs to be considered is the relationship between sample length and number of samples. This variable is not only related to categorical memory but also to temporal constraints in working memory.

Support for social interactions was not directly addressed in this first version of the software. Our group has published several studies based on music collaboration as implemented in CODES [17, 18]. Because of its conceptual simplicity, time tagging may afford easy integration with collaborative music tools in ubiquitous contexts. But it remains to be seen whether mixing will be adopted as a community-shared activity.

SUMMARY

This study showed that time tagging can be used as an interaction metaphor for mixing. Being based on a general cognitive action-perception mechanism – anchoring – time tagging stands good chances of becoming a widely applicable model for ubiquitous interaction. At a minimum, time tagging reduces computational costs by eliminating the need for visually oriented support for mixing operations. This reduction is a key requirement for deployment on low-budget mobile platforms. Thus, time tagging stands better chances of success than graphics-based metaphors when it comes to supporting music making in the cloud.
ACKNOWLEDGEMENTS

This work is partially supported by the Brazilian research funding agency CNPq (grant to Marcelo Soares Pimenta, and grants 301982/2008-2, 478092/2008-3, and 571443/2008-7 to Damián Keller). The Ubimus Group can be contacted at ubiquitousmusic@googlegroups.com.

REFERENCES


[22] Rasmussen, J., Pejtersen, A. M. and Goodstein, L. P.

