Everyday sounds: synthesis parameters and perceptual correlates

Damián Keller, Jonathan Berger
Center for Computer Research in Music and Acoustics
Stanford University, Stanford, CA 94305, USA
dkeller@sfu.ca, www.sfu.ca/~dkeller
brg@ccrma.stanford.edu, www-ccrma.stanford.edu/~brg

Abstract

Environmental sounds present a difficult problem for sound modeling because spectral and temporal cues are tightly correlated. These sounds form classes that cannot be handled by traditional synthesis methods. Micro-level representations provide ways to control spectral and spatial cues in sound synthesis and meso-level representations determine the temporal structure of sound events. By constraining the synthesis parameter space to ecologically meaningful ranges and defining parametric transformations along perceptually relevant dimensions we are able to model sound events at the micro and meso level. The integration of these approaches into a coherent data structure extends the parameter space of ecological models to the domain of spectral and spatial cues.

Introduction

Environmental sounds are produced by the action of agents upon objects in a given space. This interaction, when constrained to a finite temporal interval, constitutes an event. The observation window of ecologically meaningful events encompasses a complete cycle of energy input and energy dissipation which corresponds to a sound with well-formed attack and decay. Events are constrained by the properties of the vibrating bodies, by the behaviors of the excitation agents, and by the acoustics of the surrounding space.

This paper proposes an algorithmic representation that defines sound events by multilevel parameter manipulation (Keller & Silva, 1995). Specifically, we address issues relevant to the synthesis of micro and meso-temporal sonic features. The discussion focuses on two areas: (1) micro-temporal and spectral cues, which are related to the types of materials and the behavior of the excitation processes, and (2) spatial cues, which result from the acoustic characteristics of the space where the events take place.

The first section of this paper discusses experiments in timbre classification which study the perceptual dimensions used in instrumental sound recognition tasks. These studies suggest that when listening to orchestral instruments listeners tune onto spectral and temporal characteristics, more specifically brightness and attack bite. The samples used for these timbre experiments fall within impulse and continuous sonic classes. Everyday sounds present more varied temporal profiles, including iterated and heterogeneous sounds. The following section addresses other sonic factors, such as granularity and affordance, which might play a relevant role in the recognition of environmental sounds. Our revision of perceptual cues utilized when listening to environmental sounds is completed with a section on the impact of the environment's acoustics on sound events.

The last part of this study describes an algorithmic representation of ecologically-based events, focusing on parameters related to spectral and spatial cues. The section on properties of coincidences provides a formalization of simultaneous independent random processes and analyzes

how these processes can be constrained for their use in ecologically-based synthesis. The subsequent section proposes an algorithm for generation of events based on interactions between excitation and damping. The paper concludes with a discussion of the relevance of the proposed representation in ecological modeling. The most important aspects of this formalization are the ability to deal with correlated processes and the introduction of state-dependent dynamics.

Timbre space: its limitations

Environmental sounds have only recently begun to receive more attention from researchers in various fields (Gray et al., 2001). On the contrary, instrumental sounds have been extensively studied by musicians, psychologists, physicists, etc. Although research on instrumental timbre is not always relevant to everyday sounds, aspects related to spectral and micro-temporal characteristics can be safely extended to this field. A classical example of the methods used in studies of timbre is provided by Grey's (1977) instrumental timbre experiments. Grey recorded and analyzed 16 notes from 12 orchestral instruments. He used isolated sounds with the same pitch (Eb above middle C) and durations ranging from 300 to 400 ms. These sounds were analyzed using heterodyne filtering (i.e., Fourier-based analysis) (Hourdin, Charbonneau, & Moussa, 1997, 55) to extract their component frequencies, to eliminate micro-variations in the components and to equate them for pitch, loudness and subjective duration. Grey presented pairs of tones to 20 experienced musicians to obtain similarity ratings on an arbitrary scale. By using a multidimensional scaling procedure (Kruskal & Wish, 1978) on the averaged similarity judgments, he found the best configuration of tones so that the distance between the sounds was correlated to their similarity. This statistical analysis suggested that the listeners were basing their judgments on three dimensions.

Grey examined the time-spectrum analysis of the sounds to find what physical dimensions were related to the three perceptual dimensions found. The first dimension corresponded to the spectral energy distribution of the sound, which correlates to brightness (Wessel, 1979). This dimension can be measured by using the amplitude-weighted average of the components, or centroid (Iverson, 1995). The second dimension is related to two factors: the synchrony of onset and decay between the harmonics and the spectral flux, i.e., the coherence among the micro-variations of the components (Krumhansl, 1989; McAdams, 1993). This dimension can be associated to static quality (e.g., woodwinds) versus dynamic quality (e.g., brass) (Handel, 1995). Finally, the third dimension reflects the temporal characteristics of the attack which are determined by the 'bite' or 'explosiveness' of the sound. This feature can be quantified by measuring the synchronicity and harmonicity among component onsets for various regions of the spectrum.

Timbre mapping onto Euclidean distances requires several theoretical assumptions: (1) timbre space dimensions should be orthogonal, i.e., changes on one dimension should not affect the other dimensions; (2) the distance between samples A and B has to be equivalent to the distance between samples B and A; (3) the topography of the timbre space must be independent from the number and type of classes it contains. The first issue - interaction among sonic variables - has been addressed in several studies (Tróccoli & Keller, 1996). These studies concluded that sonic changes yield co-determined perceptual variables. Handel (1995, 441) suggests that "the cues for timbre depend on context: the duration, intensity, and frequency of the notes, the set of comparison sounds, the task, and the experience of the subjects."
Grey's (1975) study throws light upon the second problem arising from timbre mapping, that is, the equidistance between two points in the space. Grey presented a sequence of 11 tones where the first and last ones were two instrumental timbres and the other eight were progressive spectral interpolations between these two tones. The subjects were asked to report where in the sequence the timbre changed from the first tone to the second tone. The responses of all subjects tended to 'stick' to the first tone switching to the second timbre only after the dimensional midpoint was exceeded. Grey (1975, 82) reports that this effect occurs within a range of 25 to 45 percent of the

extent of the interpolations. Here, the order of presentation of the stimuli plays the main role, i.e., if spectrum A is morphed into spectrum B, the perceptual system stays longer within A's region than within B's. If spectrum B is presented first, then hysteresis occurs in the opposite direction. In other words, the transformation is asymmetrical.

Regarding the third issue, McAdams (1999) addresses the topography of the timbre space and its relationship to the number and type of timbre classes. McAdams et al. (1995) propose the use of specificities to obtain dimensions that correlate more clearly to the acoustic parameters. Specificities are the distinguishing features of a sonic class, such as the hollowness of the clarinet sound, the noise of the harpsichord pluck mechanism, etc. These characteristics are not accounted for by the dimensions of the timbre space. Paradoxically, these cues may play a key role in the recognition of environmental sources. The interaction between a source of excitation and a resonant body provides information on the affordances of the object and on the behavior of the agency (Keller, 1999b), thus reducing drastically the classes available in recognition tasks.

As we have previously discussed (Keller, 2000a), the types of samples used in timbre studies bias the results obtained. In Grey's (1975) study, the use of a single instrumental sound sample implies a homogeneity within its class: all sounds in this class are perceptually equivalent and any interactions among classes or among contexts are precluded. In Grey and Gordon's (1978) experiments, timbre is assumed to be constant across the whole instrumental range. Variations in intensity are excluded and the analyzed sounds are restricted to the middle range of the instrument. Other factors overlooked in timbre space studies are spatial placement and source movement. Recent findings in psychoacoustics stress the importance of dynamic cues in auditory placement of sound sources (Grantham, 1995). Changes in static and dynamic variables of the source signal and differences in phase among reflections have a striking impact on the characteristics of the resulting timbre. The fact that the space where sound sources occur produces temporal and spectral effects is usually not addressed in timbre studies. Sounds are always played within a homogeneous, everlasting space.

Regarding the categorization of environmental sounds, a major limitation of the studies discussed thus far is their systematic exclusion of sonic classes outside the instrumental palette. Therefore, the timbre dimensions obtained are only representative of impulse and continuous sounds (Keller, 1999b). Iterated sounds and heterogeneous events are not represented in the sample database. Among the environmentally relevant characteristics excluded, we can mention: granularity, elasticity, viscosity, and object state.

Granularity has been pointed out as a significant factor for categorization of several sound classes. Bregman (1990, 119) suspects that "grain analyzers in audition describe the texture of sound as having acoustic grains of some particular range of properties and that scene analysis might segregate regions of sound for which the descriptions were different enough. This segregation might be expected to occur along the two dimensions of the spectrogram. Along the time dimension, the process might segregate different sequential events, leading to the perception of event boundaries. In the frequency dimension, different frequency bands with the same granular properties might be grouped together to create the spectral profile of a single source of sound." We have conducted informal tests which pointed to, at least, three factors that influence the segregation among granular streams: location (in the stereo field), timbre, and density - see sound examples in (Keller, 2000d). Di Scipio (2000) has used iterated functions to synthesize granular textures. Iterated functions present the advantage of being compact representations that exhibit very complex behaviors. Their drawback is the difficulty to predict their sonic output. Miranda (1995) has employed cellular automata to control banks of oscillators for granular synthesis. In a recent paper (Correa, Miranda, & Wright, 2001), he proposes a categorization of synthetic granular sounds comprising five classes: fixed mass, flow, chaotic, explosive, and general textures. To date, no systematic work has been done to establish a perceptual classification of granular sounds.

Bregman (1990, 120) underlines the speculative stage of work in perception of granular sounds. Nevertheless, he recommends possible strategies for the organization of grains: (1) synchronization at the onset, (2) time-varying temporal density, (3) and statistical distribution, the average of variations for a given parameter would allow to differentiate a sequence of textures. In relation to the latter issue he argues that extended exposure is necessary for the perceptual system to settle on a particular percept. Especially relevant is his suggestion that roughness (of surfaces) could be defined by the following granular synthesis parameters: density, grain duration, and grain envelope (Bregman, 1990, 118). The effects of grain envelope, or grain window transformation have been discussed in (Keller & Rolfe, 1998). The section "Computational Representation" presents a further refinement of these parameters. Grain distributions are defined both by the characteristics of independent streams and by their global effect. The spectral and spatial cues are controlled at a local level by the type of samples included in the database and by spectro-temporal transformations; the global parameters are phase-synchronicity and overlap among streams.

Table 1 summarizes the perceptual dimensions utilized in the recognition of instrumental classes and environmental sounds. Brightness is especially relevant for continuous instrumental sounds. It is calculated as the amplitude-weighed average of all partials for the whole duration of the event. Attack bite plays an essential role in the perceptual classification of impulse sounds. It is measured as the logarithm of the rise time: from a minimum threshold to the time when maximum amplitude is reached (Iverson, 1995). Spectral flux, a measure of the dynamics of the spectral envelope, might be important for the classification of time-varying textures. The spectral envelope is obtained by calculating the running mean of the amplitudes of every three adjacent harmonics (McAdams, 1999). By taking the logarithm of the standard deviation of component amplitudes from this envelope we get a quantification of the spectral flux.

Table 1. Timbre parameters and acoustical correlates.

Spectral		**Temporal**		**Spectro-temporal**		**Spatial**	
brightness	averaged centroid	attack bite	synchronicity	spectral flux	deviation from GSE[1]	location	azimuth / elevation
pitchedness	harmonicity	damping / decay	amplitude envelope	spectral contour	Δbrightness centroid variability	diffuseness	spread among locations
homogene-ity	granular spectra	density / volume	granular overlap	specificities	. . .	direction / movement	Δlocation
		roughness	granular periodicity				

Affordances

From the perspective of ecological psychology, the interaction between an individual and his environment is shaped by a process of mutual determination where activity is guided by survival needs. "A species evolves to deal with its environment in ways that will ensure survival. Similarly, an individual animal learns to deal with its particular environment. These adaptations involve both a selection for certain anatomical attributes compatible with the environment and an increased sensitivity to relevant aspects of the environment. Species become physically and perceptually attuned to their environment through evolution." (Michaels & Carello, 1981, 47).

"For perception to be valuable, it must be manifested in appropriate and effective actions on the environment. And, in turn, for actions to be appropriate and effective they must be constrained by accurate perception of the environment. (. . .) The action system (effectivity structure) and the environment (affordance structure) are in a relationship of mutual constraint." (Michaels & Carello, 1981, 54). This affordance structure is determined by the characteristics of the objects that surround us. Thus, a glass can be grabbed, filled, hit, rubbed, or broken; but it cannot be bent. A steel pipe can be bent but it does not afford crashing. A rubber ball bounces but a water drop does not. The constraints established by the affordance structures provide key information about the properties of the materials, the shape of the objects, and the type of excitation exerted on them. The way this information is employed depends on the goal of the agent and it is specific to the forms of interaction established by a particular agent with a particular object. “Information is depicted to some degree as 'personal,' as opposed to a detached list of qualities that could serve all organisms equally well.” (Michaels & Carello, 1981, 45). Taking into account these constraints, we have proposed that the ongoing interaction between the individual and the environment establishes a highly individual dynamic process, i.e., the personal environment (Keller, 2001).

The concept of affordance can be mapped directly onto the algorithmic processes used in ecological modeling. To obtain cues which are consistent with the actions and objects of our everyday environment, excitation patterns have to be constrained to Earth-bound behaviors. These behaviors are shaped by the interaction of energy-input and energy-dissipation processes. The manner in which an object is excited determines the attack’s acoustic characteristics. Its perceptual correlate has been described as the sound’s ‘bite’ (Grey, 1977). This attribute can be molded by controlling the phase-synchronicity of granular streams. Exact alignment produces the sharpest attacks and sluggish attacks can be synthesized by increasing the offset among streams. As with other ecologically-based processes, the parameter ranges are limited by the properties of the exciting and the resonant objects. For example, impacts produced by compound objects are modeled by multiple, phase-asynchronous streams.

An important issue in implementing ecologically-based models is the consistency of the sonic event’s spectral and meso-temporal characteristics. The micro-temporal level is directly related to the affordance structure of the object, including its material and shape. The meso-temporal patterns provide cues related to the excitation processes to which the object is subjected and to the state of the object. Clearly, the classes of objects that afford breaking will feature spectral characteristics consistent with ‘breakable’ materials. For example, a glass bottle will produce a spectrum identifiable as glass and will generate a breaking meso-temporal pattern when it changes state. Similarly, the action of pouring will only be consistent with spectral and micro-temporal behaviors produced by fluids. Therefore, in the synthesis process the spectral classes constrain the meso-temporal patterns available, and simultaneously the temporal structures determine the possible spectral configurations, thus restricting the parameter space.

Events in space

Everyday sound events occur in varied surroundings which modify their temporal and spectral characteristics. "The environment destroys any simple invariance between an acoustic event and the waveform that arrives at the ear. (. . .) The perceptual system thus faces a general problem of separating properties of the source from the properties of the [environment]." (Darwin, 1990, 220-221). The perceptual cues provided by the surroundings allow for localization of sound sources, and estimation of movement and direction. “To a first approximation, the perceived location of an auditory object in three dimensions is governed independently by subsystems that determine its horizontal position based on interaural differences cues, its vertical position [related to] pinnae-based cues, and its distance by a constellation of cues including intensity, reverberation, and spectral content.” (Grantham, 1995, 338).

Rasch and Plomp (1982, 144) review some factors that influence our perception of indirect sound, that is, sound reflected from the surfaces that surround the listener and the sound source. Regarding subjective evaluation of indirect sound, they provide the following guidelines: (1) Indirect sound arriving within 25 ms. after the direct sound counts as direct sound. (2) Sound arriving between 25 and 80 ms. after the direct sound is divided into: (a) sound arriving at 40 ms. or less is counted as direct sound, and (b) arriving from sides and rear is indirect sound. (3) Sound arriving later than 80 ms. is classified as indirect sound.

As far as lateral reflections go, Rasch and Plomp (1982, 145) report several time constraints. (1) Reflections within 10 ms. result in a subjective sidewards shift of the sound source. (2) Strong reflections later than 50 ms. are perceived as echoes distinct from the direct and early indirect sound. (3) Reflections 20 to 25 dB weaker than direct sound are below detection threshold. (4) Reflections between 40 and 100 ms. are responsible for the spatial impression. Contrastingly, reflections between 10 and 40 ms. are added to the direct sound, distorting its timbre.

Reverberation provides spatial cues but simultaneously blurs changes in the amplitude envelope of the source signal. Thus, the hearing system has developed several mechanisms to separate sources from reflections. “Through the operation of a precedence mechanism, the apparent positions of transient sound sources are not much affected by room reflections. (. . .) The echo-suppression aspect of the precedence effect operates within a long-term temporal context (measured in seconds), in which its effectiveness builds up or is released.” (Grantham, 1995, 339). "Listening with two ears rather than one overcomes some of the effects of reverberation. Signal processing techniques for de-reverberation exploit the fact that the reverberant part of the waveform is substantially uncorrelated." (Darwin, 1990). Another form of de-reverberation is carried out by using previous context to estimate the difference in duration between the source and the reverberant signal. All in all, the perceptual system attains a fairly robust separation between source sound and reflections.

An issue that has not been given enough attention “is the importance of *dynamic* cues in distance perception. (. . .) Recent data suggest that *changes* in sound pressure as an organism moves (or as the target moves) might provide information beyond that available when the target and observer are static.” (Grantham, 1995, 328). These results give further support to the integration of individual-environment processes in understanding environmental sounds and their perception. In this case, the active exploration of the environment provides a more accurate assessment of the position of the source than mere passive hearing. By the same token, head movements supply a more thorough sampling of the space since they allow for comparisons among the information obtained at the different head positions. Consequently, Truax’s (1984) concept of *active listening* could be taken a step further by proposing *listening-in-action*, in which the listener adapts dynamically his relationship to the source in order to disambiguate the perceptual information.

Computational representation

We have developed a simple computational representation for the synthesis of ecologically constrained sound events. Sound events are synthesized by controlling the distribution of granular samples through temporal patterns molded after excitation and damping processes. The source sounds consist of sample pools extracted from prerecorded sounds (Keller & Truax, 1998). The excitation patterns are produced by a multidimensional parameter matrix. The first dimension of the matrix represents absolute time, the second is an array of streams, and the third corresponds to granular parameter values. The temporal granular distributions are modeled after the behavior of real-world processes, such as bouncing, breaking, pouring, etc.

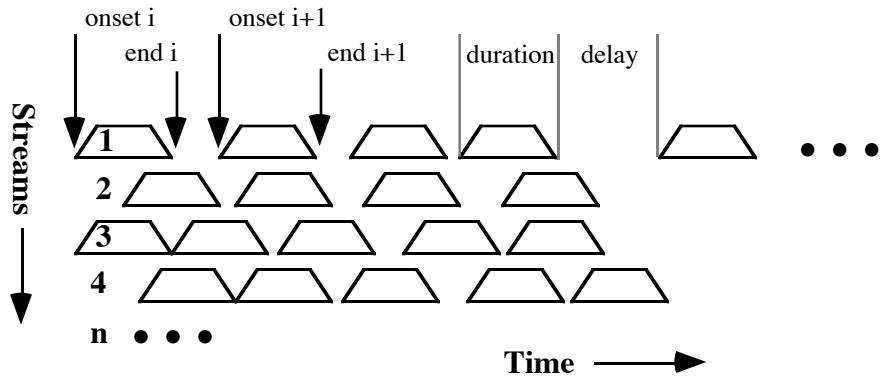

Figure 1. Parametric representation of ecologically-based events.

From an implementation perspective, the sound event is shaped by three processes: excitation, resonance and damping. In this context, the temporal distribution of granular samples is a key parameter in shaping attack transients. Resonance parameters are related to the spectral profile of the sound event. This profile can be molded by means of resonators or by controlling the phase synchronicity among granular streams. Finally, attenuation processes define how energy is dissipated. Through damping coefficients, these processes control the distribution of grains and the decay rate of spectral components.

Properties of coincidences

Given that ecological models make use of stochastic distributions, it is useful to study the properties of random processes and their possible configurations. Dziech's (1993) research suggests that a key aspect of several independent, simultaneous, random processes is the temporal relationship among the events, in other words, their coincidences. In granular synthesis parlance, this is called the overlap (Jones & Parks, 1988). Dziech studies the behaviors of random streams by analyzing the distributions of isolated streams and of several simultaneous streams. For a single independent stream, the most relevant characterization is its filling factor, or the amount of time during which granular events are active. It is calculated by multiplying the average number of onsets times the average grain duration. The filling factor of coincidence streams is equal to the sum of coincidences of all streams (Dziech, 1993).

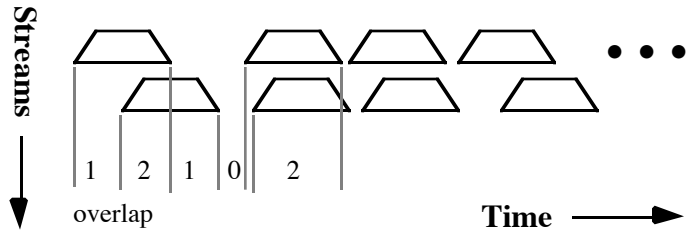

Figure 2. Overlap or coincidence among streams.

The number of coincidences among granular events within a given time frame is related to the perceptual effect of depth or 'volume' of the event (Truax, 1992). The coincidence stream is obtained by calculating the overlap among all streams. Given independent, random processes, it is equal to or less than the minimum filling factor. The amount of overlap can be increased by reducing the variance of the grain onsets and reducing the offset among streams, or simply by

increasing the filling factors.

Dziech (1993, 157) has observed that the proportion between the average grain duration of the coincidence stream and the number of streams can be described by an exponential curve. This corresponds to our empirical observation that an exponential increase in the number of streams produces an approximately constant increase in the perceived sound depth (Rolfe & Keller, 2000).

Properties of the event

The temporal evolution of a sound event is defined by dynamic interactions between two processes: excitation and damping. A simple algorithm accounts for the possible configurations of these interactions: **x(t) = x(t-1) + a**, where **x(t)** stands for the energy state of the system at time **t**, and **a** is the excitation or damping coefficient (depending on its sign). This process establishes temporal constraints on the parameter range of the excitation pattern. In other words, every event starts from zero energy and builds up at an ecologically-bound rate, until the energy input stops. At this point, the damping process kicks in reducing the energy level until zero is reached. Three dynamic evolutions are possible: (1) excitation, when **a** is positive, (2) equilibrium, when **a** is zero, and (3) damping, when **a** is negative. The excitation process shapes the event's attack dynamics. The state of equilibrium corresponds to a sustained energy input compensated by approximately equal energy dissipation. The damping process defines the event's decay characteristics. By means of a single control parameter, this algorithmic structure generates ecologically constrained meso patterns. As in real-world situations, sound events change state relatively slowly and decay gracefully.

The event's global spectral profile is defined at the microlevel by the types of samples employed in the synthesis engine. Homogeneous events are obtained by using a single class of granular material, e.g., water drops. Heterogeneous events are synthesized by means of class mixture or hybridization, or by applying transformational processes to the collected samples (Keller, 2000b). The spectral characteristics of the granular samples can be modified by using resonators, such as string and pipe models (Keller, 1999a; Smith, 1997).

Spatial cues are established by three factors: (1) local reverberation which can be implemented by convolving the granular samples with various impulse responses, (2) phase-controlled granular processes in which the phase offset and the amplitude scaling provide cues akin to early reflections or sonic 'volume' (Truax, 1992), and (3) localization, or source placement within the stereo ou multichannel field, which is implemented by means of amplitude panning and loudspeaker-dependent delays (Lopez-Lezcano, 2001). Furthermore, the dynamic manipulation of location provides directional cues for source movement and diffuseness, or the event's spatial *spread*.

An illustrative example of the relationship between sonic cues and synthesis methods is supplied by the first author's piece *IQ²*. This video and sound installation was presented at the *Iron Works Gallery* in Vancouver during September 2000 (Keller & Knox, 2000). The sonic material consisted of two layers of events exploring hybrids of water drops and metallic impulses. One layer was constantly present and furnished an artificial acoustic space that took advantage of the reflective properties of the gallery. The sounds in this layer were created through convolution of reverberant water drops and impacts on metallic surfaces. Sounds were laid out on two tracks routed independently to two sets of speakers. These speakers were placed on the floor pointing to the ceiling. Wall and ceiling reflections reinforced the reverberant characteristics of the soundtrack. Another layer was diffused through a group of four speakers hanging from the ceiling. This layer featured several types of events synthesized along a continuum that went from harsh metallic impulses to smooth water stream sounds. The events were activated by five motion sensors placed on the walls of the room. The quantity and behavior of the spectators determined what classes of events were triggered.

Discussion

A high-level representation for synthesis of excitation patterns presents several advantages over other control methods. The multidimensional matrix provides a general way to deal with multistream granular textures and grants direct access to perceptually relevant parameters: (1) phase-correlation, (2) time-dependent overlap, and (3) time-dependent synchronicity (Rolfe & Keller, 2000). These parameters control perceptual timbre dimensions, such as attack 'bite', decay rate, and sonic depth or volume. By combining synthesis methods with diffusion parameters, this data structure will provide a handle on spatial cues such as location, diffuseness, and reverberation. Unlike asynchronous granular synthesis models (Roads, 1996), a matrix-based representation can easily incorporate correlation among variables and state-dependent processes.

Although rate of change is a major factor in the organization of auditory scenes, correlation among variables also seems to play an important role. Even within the 'mess' of fast changing, widely varying multiple sources, organizational *strategies* allow us to pick up relevant cues. Phase coherence within sonic textures, similar rate and type of spectral changes, and correlated delayed signals, all point to a single energy source. It is not unlikely that the auditory system applies similar comparison strategies to completely different functional tasks, such as separation of source and reflected sound, extraction of resonant characteristics of a vibrating body, and estimation of the number of similar sources.

The relevance of synthesis processes linked to the state of the sound source becomes clear when interactions among multiple sources are considered. Environmental events occur in the context of extremely complex backgrounds. When biological agents are involved, the characteristics of the events are dynamically modified by the action of the agent on the environment and by the state of the environment during the agent-object interaction. A representation which updates its state as new information becomes available allows for environment-dependent changes of the event's characteristics.

Conclusion

The representation for ecologically-based sound event synthesis proposed in this paper brings together two control levels: micro and meso. Because these levels are handled independently, this framework can be applied specifically to the resynthesis of environmental sounds and generally to other types of sound processing. This framework encompasses spectral, meso-temporal, and spatial cues. Events are shaped by means of excitation and damping processes that determine the attack and decay characteristics. The excitation-damping process facilitates the synthesis of ecologically constrained sound events by establishing meso-level patterns from a single parametric control. Through multilevel algorithmic structures, attack and decay transient dynamics, and sound texture can be controlled in detail.

Future work will extend this framework to account for interactions among several sound sources. Data needs to be gathered on how events are distributed at different geographic locations during long-term spans. By using large temporal windows we expect to observe high-level temporal patterns. The match between information on the social factors at play (Keller, 2000d) and the analysis of long-term sonic processes, should provide more effective methods to model the dynamics of our sonic environment.

References

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Correa, J., Miranda, E., & Wright, J. (2001). Categorising complex dynamic sounds. *Electronic Musicological Review, 6*.

Darwin, C. J. (1990). Environmental influences on speech perception. In W. A. Ainsworth (Ed.), *Advances in Speech, Hearing and Language Processing* (Vol. 1, pp. 219-241). London: JAI Press.

Di Scipio, A. (2000). *Ecological modelling of textural sound events by iterated nonlinear functions*. Proceedings of the XIII Colloquium on Musical Informatics, L'Aquila.

Dziech, A. (1993). *Random Pulse Streams and their Applications*. Warszawa: Elsevier.

Grantham, D. W. (1995). Spatial hearing and related phenomena. In B. C. J. Moore (Ed.), *Hearing* (pp. 297-345). San Diego, CA: Academic Press.

Gray, P. M., Krause, B., Atema, J., Payne, R., Krumhansl, C., & Baptista, L. (2001). The music of nature and the nature of music. *Science, 291*(5501), 52-54.

Grey, J. M. (1975). *An Exploration of Musical Timbre*. PhD Thesis in Psychology, Stanford University, Stanford, CA.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America, 61*, 1270-1277.

Grey, J. M., & Gordon, J. W. (1978). Perceptual effects of the spectral modifications of musical timbres. *Journal of the Acoustical Society of America, 61*, 1270-1277.

Handel, S. (1995). Timbre perception and auditory object identification. In B. C. J. Moore (Ed.), *Hearing*. New York, NY: Academic Press.

Hourdin, C., Charbonneau, G., & Moussa, T. (1997). A multidimensional scaling analysis of musical instruments' time-varying spectra. *Computer Music Journal, 21*(2), 40-55.

Iverson, P. (1995). Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception and Performance, 21*(4), 751-763.

Jones, D., & Parks, T. (1988). Generation and combination of grains for music synthesis. *Computer Music Journal, 12*(2), 27-34.

Keller, D. (1999a). touch'n'go / toco y me voy . Burnaby, BC: earsay productions.

Keller, D. (1999b). ***touch'n'go:*** *Ecological Models in Composition*. Master of Fine Arts Thesis, Simon Fraser University, Burnaby, BC.

Keller, D. (2000a). Beyond timbre space: Hearing Seminar [Oral Communication], CCRMA, Stanford University. March 3.

Keller, D. (2000b). Compositional processes from an ecological perspective. *Leonardo Music Journal, 10*, 55-60.

Keller, D. (2000c). *Social and perceptual dynamics in ecologically-based composition*. Proceedings of the VII Brazilian Symposium of Computer Music, Curitiba, PN.

Keller, D. (2000d). *Streaming by timbre and location: sound examples*, [Web Page]. Available: http://www-ccrma.stanford.edu/~dkeller.

Keller, D. (2001). Social and perceptual dynamics in ecologically-based composition. *Electronic Musicological Review, 6*.

Keller, D., & Capasso, D. (2000). Social and perceptual processes in the installation *The Trade*. *Organised Sound*, 5(2), 85-94.

Keller, D., & Knox, U. (2000). *IQ²* [Multimedia Installation]. Vancouver, BC: Western Front.

Keller, D., & Rolfe, D. (1998). *The Corner Effect*. Proceedings of the XII Colloquium of Musical Informatics, Gorizia, Italy.

Keller, D., & Silva, C. (1995). *Theoretical outline of a hybrid musical system.* Proceedings of the II Brazilian Symposium on Computer Music, Canela, RS.

Keller, D., & Truax, B. (1998). *Ecologically-based granular synthesis*. Proceedings of the International Computer Music Conference, Ann Arbor, IL.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielsen & O. Olsson (Eds.), *Structure and Perception of Electroacoustic Sound and Music* (pp. 43-53). Amsterdam: Elsevier.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional Scaling*. Beverly Hills, CA: Sage Publications.

Lopez-Lezcano, F. (2001). *Dlocsig: a dynamic spatial location unit for CLM-2,* [Web Page]. Available: http://www-ccrma.stanford.edu/~nando.

McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in Sound* (pp. 146-198). Oxford: Oxford University Press.

McAdams, S. (1999). Perspectives on the contribution of timbre to musical structure. *Computer Music Journal, 23*(3), 85-102.

Michaels, C. F., & Carello, C. (1981). *Direct Perception.* Englewood Cliffs, NJ: Prentice-Hall.

Miranda, E. R. (1995). *Chaosynth: Um sistema que utiliza um automato celular para sintetizar partículas sônicas*. Proceedings of the II Brazilian Symposium of Computer Music, Canela, RS.

Rasch, R. A., & Plomp, R. (1982). The listener and the acoustic environment, *The Psychology of Music* (pp. 135-147). Orlando, FL: Academic Press.

Roads, C. (1996). *The Computer Music Tutorial*. Cambridge, MA: MIT Press.

Rolfe, C., & Keller, D. (2000). *De-correlation as a by-product of granular synthesis*. Proceedings of the XIII Colloquium on Musical Informatics, L'Aquila, Italy.

Smith, J. O. (1997). Acoustic modeling using digital waveguides. In C. C.Roads, S. T. Pope, A. Piccialli, & G. DePoli (Eds.), *Musical Signal Processing* (pp. 221-263). Lisse: Swets & Zeitlinger.

Tróccoli, B. T., & Keller, D. (1996). *A série harmônica: interpretação inter-disciplinar - física,*

psico-acústica, histórica - e o processo audio-perceptivo (Technical Report 0222). Brasília, DF: Fundação de Apoio à Pesquisa do D.F.

Truax, B. (1984). *Acoustic Communication*. Norwood, NJ: Ablex Publishing Corporation.

Truax, B. (1988). Real-time granular synthesis with a digital signal processor. *Computer Music Journal, 12*(2), 14-26.

Truax, B. (1992). Composing with time-shifted environmental sound. *Leonardo Music Journal, 2*(1), 37-40.

Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal, 3*(2), 45-52.

Appendix

Table 2. Parametric model of streams for ecologically-based sound synthesis.

parameter	description	computation
onset i	beginning of grain	temporal distribution
end i	end of grain	
dur i	grain duration	filling factor
amp i	grain amplitude	
env i	grain amplitude envelope (window)	spectral profile
sample i	grain sound file	sample pool
channel i	source sound file channel	stereo placement
onset(i+1)	beginning of next grain	
del i	delay between end i and onset(i+1)	Δonset i - dur i
Δonset i	delay between contiguous grains	onset(i+1) - onset i
overlap(t)	number of active grains at time t	∑ grains(t)
Δphase i	phase-offset for grain i between stream n and the contiguous stream	onset i(n) - onset i(n+1)
onset s	beginning of stream	attack synchronicity
sample s	sample pool for stream	sound class
path s	structure for spatial placement of stream	spatial localization
energy(t)	energy level at time t	excite(t) - damp(t)
damp(t)	damping coefficient	
excite(t)	energy input coefficient	

1. GSE: global spectral envelope.