INTERDISCIPLINARY SCIENCE REVIEWS
2022, AHEAD-OF-PRINT, 1-15
https://doi.org/10.1080/03080188.2022.2035103

Speech, music, soundscape and listening: interdisciplinary explorations

Barry Truax

School of Communication, Simon Fraser University, Vancouver, Canada

ABSTRACT

The author discusses the relationship between experiential listening knowledge and scientific interdisciplinary knowledge in regard to sound, with particular emphasis on soundscape composition and electroacoustic signal processing.

KEYWORDS

Listening, soundscape, electroacoustic composition, microsound, music, speech, environmental acoustics, psychoacoustics

CONTACT Barry Truax truax@sfu.ca School of Communication, Simon Fraser University, 8888 University Drive, Burnaby BC Canada V5A 1 S6

1. Introduction

Sound has traditionally been studied and knowledge about it practiced in the specific areas of speech, music and the sonic environment. The sciences of acoustics and psychoacoustics have contributed a significant knowledge base for these areas based on a traditional energy transfer model and response characteristics. However, over the last century, audio technology has transformed our relation to these acoustic sources by making all sounds available for creative production, as well as their commodification. Practices such as electronic music, acousmatic music, text-sound and soundscape composition, among others, have enlarged the scope of music to the point where alternative terms are needed, such as organized sound and sounding art. With this expansion comes a need for an equally expanded interdisciplinary knowledge base, which pedagogical training has been slow to formulate.

The experience of creatively working with sound directly, as in the electroacoustic studio with audio signals – that I will refer to here as ‘sound materials’ – demonstrates that *listening* takes a heightened and central role, providing its own form of sensory knowledge that may guide the process. Technical knowledge about the effects of various types of audio manipulation is also necessary in order to ensure efficiency in the work involved, but most practitioners would likely agree that listening to the results is the best way to evaluate the effects of such processing. In the traditional analog studio, many of the typical processes were guided by interactive ‘hand-ear’ co-ordination, that is, by varying the parameters of a process in real time and assessing the aural effects, a kind of ‘fine tuning’ guided by listening. Many equivalent digital processes are more algorithmic in the sense that such parameters are pre-specified before they are implemented, and in some cases, there is a delay time to allow calculations to be made. With faster processing speeds, these delays may be shortened or even become negligible, but not surprisingly, use of 'presets' of parameter values may be encouraged by the software, with a decreasing likelihood that listening as a fine-tuning of the design process will occur.

What I want to emphasize in this article is the relationship between what I will call listening knowledge and interdisciplinary knowledge, in other words, the relationship between the experiential and the scientific, and specifically how contemporary electroacoustic technology allows them to interact. Traditional acoustic knowledge has been based on models of how acoustic energy is transferred from the source, through a medium to the 'receiver.' The tradition of psychoacoustics, based in nineteenth-century studies of psychophysics, picks up the train of energy transfers which it regards as stimuli, and then determines what the subjective response is for various human and non-human species. With electrification in the early twentieth century, psychoacoustics devised an experimental methodology based on audio signals whose individual parameters could be varied independently, thereby allowing response patterns to be quantified and mapped for parameter interaction, for instance, how subjective loudness varies according to the frequency and amplitude of the signal. Not surprisingly, audiology was brought into the same methodology such that hearing ability could be normalized and measured similarly.

Because listening clearly involves cognitive processing, it has been much more challenging to research, though considerable progress has been made. For the purposes of this article, I will define listening as the processing of the sonic environment to extract usable information for the brain which can influence human behaviour. It should be noted, in the spirit of 'aural diversity,' that the listening process may be very different for each individual depending on how their hearing ability functions. But whatever the actual process, I want to focus on how listening involves complexity. For instance, how do we separate complex patterns of acoustic energy into the recognition of multiple sources, and what are our limits for doing so? Or, how do we rapidly detect acoustically complex timbres and textures, group them into specific percepts, and track their evolution? And, in the case of our most basic aural orientation in an environment, how do we simultaneously process sound sources with the spatial information that is imbedded within them?

To bring this subject back to the electroacoustic studio and its affordances, I want to emphasize that what is often referred to as 'everyday listening' – with all of its actual complexity and variability as suggested above – may become heightened through our interaction with audio technology to levels of what I call 'analytical listening' (Truax 2001). The key elements are repetition that allows us to focus on details that may otherwise go unnoticed, and interaction that allows repeatable changes to be effected and evaluated. The epistemological process that is involved, which I am arguing combines listening knowledge with scientific knowledge, can then be linked to a broader contextual knowledge, as to what this aural experience means and suggests in the real world, not to mention extensions into the imaginary.

2. Speech, music and soundscape

Whether the subject matter is speech, music or environmental sound, the role of listening changes the focus from energy or signal processing to perceptual experience and communication (Truax 2001). We see, for instance, how long it has taken for automated speech recognition to develop to its current state, and with what vocabulary and semantic limitations. And yet, we typically can recognize both the semantic content of language and its paralinguistic features, that is, the analog form of the communication in terms of pitch inflections, loudness contours, rhythm, articulation, non-verbal elements, and the use of silence. Moreover, even without particularly attentive listening, we can surmise the emotional state of the speaker, what is intended (or unintended), as well as what might be referred as 'meta-data,' such as irony, sarcasm, deceit, joking, teasing or simply reaffirming a relationship or power differential. The subtlety and depth of such communicational interpretation indicate the sophistication of how everyday listening can function.

Musical listening seems to be a more specialized form of multi-dimensional perception, and lies beyond our scope here, given its cultural variety and complexity. However, since our focus from here on will be the soundscape aspects of listening, we can comment on the overlap between music and soundscape, where it has become an environmental accompaniment (Truax 2011). Its role as a background stimulus originated in industry prior to the Second World War, where it was shown to enhance worker productivity, and then in the postwar period when it became the designed accompaniment of commercial consumer contexts. In the meantime, radio provided the domestic counterpart to this form of background listening, and shaped its programming structure to complement that form of ambient listening, while still integrating commercial messages into its flow. It is tempting to think of this type of 'non-listening' as an inferior form, but my contention is that it is an extension of everyday background listening which is characterized by all of the same processing and recognition strategies associated with listening – but without foreground attention being invoked. As such, I refer to it as 'distracted listening,' which is still a form of listening that the mind requires to deal with non-salient sounds in everyday situations.

The ISO Working Group 54 has recently established a definitional clarity between the objective 'sonic environment' with its quantitative measurements, and the 'soundscape' which is defined as how the sonic environment is *perceived* and *understood* by individuals in context, a distinction that was first introduced by the World Soundscape Project (WSP) at Simon Fraser University in our *Handbook for Acoustic Ecology* (Truax 1999). The Working Group is also developing criteria for soundscape data collection and reporting, with recommendations, for instance, to involve soundwalking and other *in situ* forms of qualitative listening appraisal. Soundwalking as a dedicated listening walk has been a central methodology for the WSP for the last 50 years, and has been proved to be effective in a variety of forms and contexts. Its environmental goal is to strengthen listeners' awareness and contact with the soundscape, and hence their recognition of the importance of an acoustic ecology in their personal and social lives. As a social strategy, it places the emphasis on the 'quality of life' that sound can reinforce or damage. And, as a political goal, it provides an example of an additional dimension of environmental sustainability that needs to be protected and enhanced.

From this brief overview of how listening interacts with all forms of aural communication, it should be clear that the discussion has the possibility – even necessity – of bringing a wide range of disciplines to bear on its implications, and not just acoustics, psychoacoustics and bioacoustics, but also geography, anthropology, psychology and sociology, to mention just a few. Traditionally sound has not been a focus of these social science disciplines, but it is encouraging to see that changing. A multi-disciplinary approach involving such collaborations is long overdue, particularly to tackle complex urban and environmental issues involving sound. However, we can also see various interdisciplinary approaches emerging, many under the general rubric of sound studies.

In his introduction to the Sound Studies Reader, Sterne (2012) lists several criteria to guide research in sound studies, including interdisciplinarity, reflexivity, historicity and criticality. To me this suggests that we cannot just bring together seemingly relevant disciplines to bear on a problem or issue. Instead, we need to critically evaluate how traditional knowledge about a topic has been created historically, culturally and politically. Indeed, when R. Murray Schafer, founder of the WSP, departed from the traditions of anti-noise discourse which he had practiced for several years, and in the early 1970s suggested the positive, listener-centred focus of the soundscape, he was essentially creating an interdisciplinary concept that today is generally known as acoustic ecology (Truax 2008, 2019). It challenged the hegemony of the quantitative objectification of the acoustic environment through acoustical engineering, the laboratory-based experiments of psychoacoustics which treat sound in isolation, the visual bias of much of the humanities and social sciences – and the exclusion of environmental sounds as musical material, or more specifically, as music in and of itself.

3. Soundscape composition: environment as music

In contrast with music as environment, as discussed above, we can treat the soundscape as if it were music, and even place it within what traditionally would be regarded as a musical and artistic context. A key element in this inversion is that we listen to the soundscape with the same focus that has traditionally been applied to music, at least before it became an environmental accompaniment. But would we bring an equivalent form of aesthetic judgment to the experience? Most likely, but that does not necessarily mean that it would or should be the sole criterion for soundscape design (Truax 2011).

However, before we approach that question, let us examine how the sonic environment differs from traditional musical experience. At a basic acoustic level, most musical sounds are composed of a spectrum (i.e. frequency content) that is comprised of discrete frequency components, usually harmonically related by simple integer ratios, but in the case of metallic instruments, inharmonically related as well. These types of spectra are heard as having both pitch and a characteristic timbre or quality. In contrast, most environmental sounds do not have the periodic quality that produces pitch, except in very specific circumstances, and are more likely to be described as having a texture. Their spectra are typically described as noise-like with varying bandwidths from narrow to quite broad, as a result of random fluctuations of sound pressure.

Given that musical instruments have been developed with highly refined timbral possibilities, with the associated psychoacoustic recognition strategies to discern their complexity, we may be tempted to regard environmental sounds as less ‘orderly’ and indeed, much less research has been done concerning their perception. The ancient examples of prototypical musical instruments refer to quite specific physical configurations, such as the air column in a bone flute or conch shell, or the pitch produced by a stretched tendon fixed at both ends. From an acoustic perspective, only the one-dimensional forms of the string and air column can actually produce a harmonic spectrum composed of pure frequencies (the result of what are known as modes of vibration). In fact, the very specificity of such unique pitched sounds, as well as their relative purity compared with the sounds of nature, may account for the sense of wonder that infuses the origin myths of music across various cultures.

The other source of reference sounds is the human voice. From an acoustic perspective, its main characteristic is the ability to produce a periodic (and therefore pitched) sound by the vibration of the vocal folds, combined with a variable set of resonances in the vocal tract known as formant regions. Given the flexible shape of the vocal tract (by not being a cylindrical air column but a variable resonating space), the formant spectra it produces are broader energy regions than the narrowly defined harmonics of a tube and it is these that define a vowel. Moreover, tongue position can smoothly interpolate between various resonant configurations, an ability seldom found in inanimate environmental sound sources. Within certain physiological limits, the pitch of the vocal folds and the resonant formant regions are largely independent of each other, thereby creating two levels of perceptual distinction, which is fundamental to spoken language. Put simply, the formant spectra define the vowels, and pitch patterns define paralanguage or other semantic features in most languages.

The other key characteristic of vocal sound is the articulation function of the consonants. They are formed by air passing through the vocal tract (with and without the vibration of the vocal folds, termed voiced and unvoiced consonants respectively) that is further defined by the tongue position and the manner of articulation. Two examples of the manner of articulation are the fricatives, produced as narrow bands of noise because of the airflow (e.g. fff, sss, shh, chhh, etc.), and plosives, produced by a pressure buildup that is suddenly released (e.g. b, p, t, k, g). Both sets of these consonants have voiced and unvoiced variants. The consonants can be regarded as attack transients before a vowel, similar to how a musical instrument produces the beginning of a sound; that is, how the resonant acoustic energy is set in motion, such as a string being bowed or plucked, or an air column being activated by the embouchure of the mouth or the vibration of a reed.

This brief and simplified summary of speech acoustics is important to our argument because it leads, first, to two main psychoacoustic modes of analysis, the temporal envelope of the amplitude of a sound and its spectral envelope, the latter being a spatial distribution of resonant frequencies along the basilar membrane in the cochlea which is projected as a spatial pattern of neural impulses in the auditory cortex. In terms of speech acoustics, the primary recognition strategy is the spectral envelope for vowels, and the temporal envelope for consonants. The brain's strategy of using multiple and complementary types of analysis for rapid recognition of sound largely accounts for its efficiency in doing so. Secondly, these recognition strategies apply equally, though usually on a larger time scale, to environmental sound recognition.

Environmental sounds, other than soundmaking by other species, are generally powered by natural energy, as distinct from the human breath, and so their spectral and temporal envelopes are prolonged, even to the point of being continuous. In such extended cases, they may be thought of as gestural, which is the result of sustained but variable energy input where the pattern of variation is arguably more important than the specific spectrum of the sound source. However, impacts between objects, whether by natural or human activation, are also common in environmental sounds and tend to inform us of both the nature of the activating energy (termed the excitation function in acoustics) and the acoustic qualities of the source receiving the energy. It is remarkable how effortlessly we obtain both types of information simultaneously in such interactions. We can readily identify the extrinsic cause of the impact, and the intrinsic character of what received the energy.

Environmental sound textures often exhibit granularity, because they are composed of a myriad of smaller sub-events. Hence, there are many micro-level temporal envelopes that resemble impulses, and it is their density and 'bandwidth' that influence the overall texture. I am using the term bandwidth to indicate not only the range of frequency differences, the original meaning of the term, but also the range of any other acoustic parameter such as amplitude, duration or envelope shape. In short, we describe these textures as being stochastic, which refers to a degree of randomness at the micro level, but a perceivable density at the macro level. Think of rain falling on a surface, or wind passing through a stand of trees. Just as with the percussive impact, we hear the influence of the material being activated, as well as the overall texture that is produced through its density of smaller events. The stochastic texture, then, is often a pleasing balance of the unpredictability of the micro level combined with a predictable, if changing, textural density at the macro level.

When any such environmental interaction of energy and materials exhibits resonant qualities, similar to formants, or temporal patterns similar to bodily functions, it becomes very easy to ascribe a human character to the sound. Wind passing through a narrow gap is often described as whistling or sighing, depending on the type of interaction. Slowly breaking ocean waves can resemble relaxed breathing patterns, or when they occur faster, can be described as 'angry,' as in a storm. Similarly, other forms of human activity, such as walking or hammering represent a range of rhythms that have human associations. When mechanical or electrical energy takes over and loses the constraint of human energy patterns and limits, we regard them as other-than-human, although the possibility of textures and internal rhythms remains, sometimes reassuringly.

Horns, whistles and bells as signalling devices seem to be the most 'musical' extension of environmental sounds because of their pitch components, but given the stochastic and textural qualities of other environmental sounds, combined with species soundmaking (even if on a different temporal or pitch scale than the human equivalent), we can experience the soundscape as aesthetically pleasing, even if it has 'dissonant' noise elements at times. There will also likely be many levels of psychological associations accompanying such sounds, ranging from placid to dangerous, among other reactions. Taken together, all of these aspects of environmental sounds form a rich palette with which to work creatively.

At a macro, structural level, recent bioacoustics has made us aware that an acoustic habitat is inherent and vital to the natural world, as most clearly indicated by the Acoustic Niche Hypothesis (ANH), as documented by Krause (2012) since the 1980s. He has shown that species soundmaking occurs in non-overlapping frequency ranges, such that a clear channel of communication is available. This intricate spacing can be quite complex, for instance, when formant regions of one species fall in between those of other species, such as a narrow band of insect sound textures. Moreover, Monacchi (2016) has also identified instances, where, like musical counterpoint, species soundmaking 'takes turns,' particularly if they each fall into similar frequency bands.

Human soundscape design can learn from these natural bioacoustic examples, both in how fragile they are in terms of disruption (usually from human intrusions), and how essential they are to our orientation and optimal functioning as a community. Human hearing also provides a greater number of possible bandwidths to occupy, as measured by the critical bandwidth of the auditory system, that is, the resolving power for frequency along the basilar membrane in the cochlea, as referred to earlier. Researchers have identified about 24 such bands, which is more than for birds and mammals. When a stronger sound occupies the same band as a quieter sound, it is said to mask it, that is, make it much less audible, whereas when they occupy different bands, such as birds and traffic, both sounds remain audible, despite the power imbalance. The problem arises when most urban sounds are very broadband in nature, and therefore have the power to mask pretty much everything else.

To finish this discussion of how environmental sounds differ from traditional musical ones, or can be heard as extensions of them into a textural domain, we can consider some obvious spatial differences. My contention is that sound creates acoustic space through the sonic events that are shaped or 'coloured' by the physical space (Truax 2017). Traditionally, music performance can be understood as adapting to the acoustics of a given space, perhaps even exploiting some of its characteristics such as reverberation when the space is enclosed, or employing stronger timbres outdoors when the space is more open. When music moves to a more focused stage area, then it tends to assume that an audience is looking directly at the source, even though some types of performance may still distribute performers in different directions. Recording methods open up options for reproducing existing acoustics or simulating entirely abstract ones unrelated to the space where the musicians were recorded.

Soundscape composition, particularly when it has multiple loudspeakers available and placed around, above and even sometimes below the listener, has the ability to resemble the 360° immersion that we associate with a soundscape (Truax 2002, 2008). The auditory system, when unimpaired, has the ability to localize sound from any direction, but with greater accuracy for the frontal plane. In fact, spatial separation of sources actually allows the brain to distinguish them as distinct, an ability called 'cocktail party effect.' Unfortunately, many types of hearing impairment degrade the listener's ability to distinguish simultaneous sounds, and hence those affected tend to avoid crowded areas, and experience increased social isolation. However, in general, soundscape composition has the ability to extend even spatialized forms of musicmaking, and move towards its environmental models, and even beyond to completely imaginary soundscapes.

4. Environmental sound processing: working from within the sound

In this section, I will refer to studio-based composition with pre-recorded environmental sounds, although some aspects may apply to other creative uses of such materials. The first problem that students commonly find is that the quality of the materials they've recorded doesn't match the actual experience of those sources. Besides technical issues of audio quality, one problem here is that microphones don't necessarily respond to the surrounding ambience as does the ear. Most microphones are designed for directionality, in order to 'focus' on sounds coming from specific directions, but even omnidirectional microphones will fail to distinguish between sounds coming from, for instance, the front or back, in the manner that unimpaired binaural hearing does, where subtle timing and spectral cues give us useful information for localization.

Some recordists prefer binaural microphones for this reason, where the microphones are worn in one's ears, or embedded into an artificial head (or *kunstkopf*) with correctly modelled external ears (or pinnae), auditory canals and head contours. These reproduced sounds, which often provide a surprising sense of realism, need to be experienced by individual listeners using high-quality headphones, although many listeners report issues with getting a 'frontal image' unless it is suggested by movement in a trajectory, or other contextual expectations. Moreover, binaural recordings present difficulties for subsequent processing which will alter and degrade the aural cues needed for localization. There is also the issue of microphone movement in the original recording if a stationary listener does not adjust their orientation to it, that is, imagine they are moving and not the soundscape itself.

A second issue with recordings that lack the quality of the original environmental experience can only be attributed to the perceptual listening process itself which will likely focus on specific sources, relegating others to the background of perception, a type of practice we can call foreground listening. The auditory system has a refined system of separating early arriving signals direct from a source, from those arriving a few milliseconds later, an ability called precedence effect. The first arriving signal from a source is correlated since all frequencies travel at the same speed, whereas reflected sounds are usually randomly scattered and arrive later in an uncorrelated manner. Moreover, when sources are uniquely located in space, the cocktail party effect allows unimpaired hearing to group different sounds into auditory streams which can be attended to at will, as described above.

Given these issues, one of the main problems of using environmental sounds as source materials is that they can be quite complex mixtures of foreground, mid-ground and background elements. We have already referred to bioacoustician Bernie Krause choosing to record sonic environments as a whole, instead of the traditional practice of using sophisticated recording techniques to isolate specific sources for classificatory purposes. The World Soundscape Project treated their recordings similarly by recording sound events *in situ*, in order to give a sense of their social and environmental context. However, their range of recording techniques also included long takes of ambiences, as well as more closely miked recordings of specific sound sources. In practice, the WSP library of recordings was never intended to function as a 'sound effects' catalogue, many of which exist and can be purchased (Truax 2019).

The representational issue presented here can be regarded as the difference between a 'sound event' – with its emphasis on the totality of a soundscape – and the 'sound object' which attempts to isolate and focus on specific sources. In an acoustically controlled environment such as a studio or recording booth, it is easy to obtain a 'clean' recording of a source, but in everyday situations, this will be more difficult, but not impossible. Besides choosing a recording site with a favourable, i.e. low, ambient level, the main variable will be microphone distance, where yet another choice needs to be made, between a typical 'ear distance' or an even closer miking. Close proximity can imitate a listener's ability to focus attention on a source, relegating others to the background of attention. But it can also extend this perceptual process to an enhanced version of the sound that will likely bring additional detail into audibility.

In terms of studio processing, some of the composer's initial activity will be to select source material, edit out what is thought to be extraneous material, and in some cases, also filter out unwanted parts of the frequency spectrum. The most common situation is with wind noise on the microphone which adds a lot of low-frequency energy to the recording that can be filtered out with a high-pass filter set to an appropriate cut-off frequency, resulting in a 'rolling off' of the low-frequency spectrum. The limiting factor is the case where low-frequency environmental sound is also present, since the filter will affect its quality as well. However, sound sources with mid- and high-range frequency spectra are easily isolated, and from a listening perspective, often seem more realistic as a result, perhaps because as listeners we tend to ignore less salient features of the soundscape such as low-frequency energy that is commonly present.

At this point in a typical studio process, the composer is more likely to proceed with aesthetically informed decisions about processing than these simpler 'cleaning up' aspects of the source material. I like to refer to this next stage of processing as working from 'within' the sound. Explaining this distinction is problematic given the nature of syntax to involve a subject, an object and a transactional relation between them. Isn't everything performed in the studio imposed onto the material, shaping and modifying its character? In a literal sense, yes, every action is chosen and imposed, but some are more invasive than others when they appear to add something 'foreign' to the material. In fact, given the power of contemporary audio processing, it is actually quite easy to obliterate the original sound and transform it into something quite abstract. But to do this is to abandon, often to a large extent, the listener's contextual knowledge that allows the sound to be recognized and probed for contextual associations.

The aesthetic stance chosen by most soundscape composers is to preserve and enhance the listener's relationship to the real world and its cognition. The reasons for adopting this stance are many and probably highly varied. However, when we listen to the results, we usually can identify a continuum starting with what is generally known as the 'phonographic' approach (where phonography is thought of as a counterpart to photography) where the sound recordings are processed to a minimal or transparent degree, transparent in the sense that the listener will accept their aural realism, even if subtle manipulation has been involved (Drever 2017). For instance, audio listeners will not usually be troubled by irrational elements in a recording, such as time compression, particularly because memory itself creates a foreshortening of the temporal experience, eliminating less salient moments. Memory has very little to do with clock time, just as it has a weak resemblance to a more objective recording.

These psychological factors suggest that soundscape composers can continue to engage listeners with the 'realism' of composed soundscapes as more extensive processing is involved and is specifically heard as non-transparent. I find that it is at this point where working from 'within' the sound has its greatest benefit, by which I mean, using audio techniques that *bring out* aspects of a sound that are inherent to it. For instance, the process of equalization (EQ) can only emphasize, or de-emphasize parts of the frequency spectrum that are present. With speech, one uses a standard EQ process that emphasizes the 1–4 kHz region where important speech components (upper formants and consonants) have energy that is critical for understanding but are weak in absolute energy. This frequency range also happens to be the region where the auditory system is most sensitive – and unfortunately where noise-induced hearing loss usually is also the greatest. Therefore, using EQ can benefit a wide range of hearing abilities and make any sound more vivid or recognizable.

Something similar happens with processing using resonators that also emphasize frequencies that are already present in the sound. However, in this case, a temporal dimension is also added because of the feedback process that produces the resonance. As such, audio resonators emulate the acoustic process of resonance that can be thought of as a natural mode of amplification. Acoustic energy can be quite fragile in the sense that it does not transfer easily from a sounding body to the air or other medium. A resonant frequency represents a natural mode of vibration where the energy transfer is maximized. For instance, when a stretched string is plucked, the natural modes of vibration of the string (all harmonically related) are set in motion by the feedback of the vibration passing back and forth between the two ends, with the fundamental frequency of the vibration corresponding to twice the length of the string (and its integer divisibles, keeping in mind that frequency is inversely proportional to wavelength). A complex vibrating body such as that of a violin is prized for the richness of its many modes of vibration that colour the spectrum of the string into the rich timbre we associate with it, no matter what pitch is played. Similarly, a digital resonator can enrich even the weaker upper partials in a sound and prolong their decay through feedback.

Transpositions of a sound in pitch, which also alter its perceived timbre, can produce a result that is aurally related to the original. Through digital processing, the duration can be kept the same, but, following the analog procedure of slowing down the speed of a recording, it can also result in a change of duration. The classic analog effect of lowering the pitch by an octave and allowing its length to be doubled is a case in point. Octave intervals appear to repeat the same pitch, but with the extension of the sound's duration, it may take on a very different quality. For instance, lower resonances may appear richer, and micro-level detail becomes more aurally obvious. A textured sound may appear less dense and since many noisy environmental sounds naturally occur in different frequency ranges, the character and associations heard within the sound may also change. In general, we may say that such transformations produce an *abstracted* variation of the original, rather than rendering it completely abstract. In the language I have proposed here, we can also say that we are working within the sound to create a family of variants.

5. The microsound domain

Since we can regard environmental sound as occupying the widest range of aural experience, it may seem paradoxical to suggest that we can work with it the most effectively at the micro-level of the time domain, what is generally known as the quantum level. Engineering refers to this subject as the time–frequency domain, which has many applications other than for audible sound. The audio version of the domain is defined as dealing with signals less than 50 ms, which is where events called grains fuse into a continuous texture (Roads 1996). The other significant property of this domain is that any change in the time domain, such as grain shape or duration, results in a change in the frequency domain, as published by Dennis Gabor in the late 1940s. For instance, the shorter the grain, the broader its frequency bandwidth, such that a brief acoustic signal is heard as a broadband click.

The quantum nature of the microsound domain means that an uncertainty principle also applies, with an analogy to the equivalent relationship first determined by Heisenberg for electrons. He demonstrated that there was an inverse relation between what could be determined about the position and speed of an electron – the more precise the location, the greater the uncertainty about its speed, and vice versa. Speed, of course, is the rate of change of position, and in the acoustic world, frequency is the rate of change of phase in the time domain, hence the analogy.

The first connection of microsound to environmental sound is that a stochastic distribution of short grains (less than 50 ms) results in textures that may resemble those in the soundscape. In fact, my first work in 1986–1987 with what is called granular synthesis was a work titled *Riverrun*, composed with up to 32 individual tracks of grains, each of which could have densities between 1 and 2 K grains per second (Truax 1988). The form of the work was organized around the paradox that just as water streams of varying energy levels and volume are formed from individual droplets, so too could these synthesized sounds resemble the perceived magnitude of water textures and still be based on seemingly trivial grains. Therefore, the piece traces such a flow from individual granular 'drops' through vast cataracts, eventually calming to a metaphorical 'delta' at the end of its travel. Although the entire piece is synthesized, and individual textures would never be mistaken for actual water, the overarching metaphor of a river seems to be accepted by most listeners and helps them navigate through this new sound world.

The second connection with environmental sound in the 1990s was more transformational, in that such material could be stretched in time without any pitch change, a process called granulation of sampled sound using short, enveloped grains. Not surprisingly, this technique also seemed to work 'within' the sound, but the change in time scale resulted in significant perceptual differences. In the above section about the psychoacoustic processes involved in listening, we remarked on the complementary strategies related to information about the spectral envelope and the temporal envelope. Now, via granular time stretching, we could alter one without affecting the other, and hence the listener could focus more on the spectral evolution of the sound with a greater degree of analytical attention. Although it is technically possible to stretch pitched musical sound with few artefacts by a precise overlapping of windows (the equivalent of grains), this can only be done for relatively short degrees of stretching, beyond which non-linear modulation components are added to the sound that seem quite foreign to it.

My preference was to use stochastically organized grains when time-stretching environmental sounds such that texture was added, thereby abstracting it somewhat, but remaining within the textural domain commonly found in the soundscape. An early work such as *Pacific Fanfare* (1996) begins with a montage of various unaltered Vancouver soundmarks (a fog horn, steam whistle, airhorn, a historic cannon firing, and bells). These sounds are then stretched with the granulation technique, and instead of merely recognizing them as a familiar perceptual entity, we are given more time for reflective listening to their inner qualities. Similarly, the bells of a basilica in Quebec City are stretched and harmonized in *Basilica* (1992), such that their resonances are elongated and resemble the reverberation found inside the church.

A more common soundscape experience, that of commuting, is invoked in *Pendlerdrøm* (1997) based on recordings made in the Copenhagen train station including a local commuter train (*pendler* meaning a commuter in Danish). The piece alternates between sections of high realism in 8-channel surround format within the busy international station, while waiting for the commuter train to arrive, and then onboard, with simulated 'daydreams' (*drøm* meaning dream) in between the realistic sequences (Truax 2008). These dreamlike sections stretch selected and previously heard sounds with granulation and resonators to bring out their particular character and make them more musically defined. In the section onboard the local train, the commuter's daydream is suggested by small fragments of resonated signals and announcements returning as short loops, similar to what is known in German as an 'earworm' in memory – bits of sound that repeat in our minds. Therefore, not only is the outer experience of the soundscape simulated, but also the inner psychological world of the dream experience of a tired commuter who nods off on the train home. One can only speculate whether the experience of the piece might also carry over into subsequent real-world experience, which it often does with recordists and soundscape composers working with similar material.

Another time-frequency domain type of processing that has interesting soundscape connections is called convolution, a mathematical procedure where the spectra of two signals are multiplied together (Roads 1996). Although this sounds rather abstract, it is actually, in the first instance, a precise description of what happens when any sound source is 'coloured' by the acoustics of a physical space. The most common application of convolution is called 'impulse reverberation,' because it convolves an acoustically dry sound with the impulse response (IR) of a space, obtained by recording a broadband sound with a sharp attack (e.g. a handclap or a pistol shot) in any space, including the resulting decay. There are two types of alteration to the sound in this process, both of which occur in an acoustic situation. First, the spectrum of the original sound is altered in accordance with the response of the space in the frequency domain (for example, an acoustically 'bright' room, as the metaphor suggests, will emphasize the high frequencies in the sound, a 'warm' room will emphasize the midrange frequencies, and a 'dark' room will boost the lows). Secondly, the sound will be prolonged, just as reverberation adds an extra 'tail' to the decay of the sound. This property reflects the rule that the sum of the durations of the two sounds being convolved determines the duration of the processed sound. The result of the impulse reverb process is that the original sound appears to be located in the same acoustic space as the IR, and at the same distance the sound source was located from the microphone. The realism of the effect is due to the precise model represented in the mathematical algorithm.

When I first started working with convolution in 2002, I added an additional variation on the IR process by convolving the sound with itself first, and then with the IR of the chosen space. The results formed the basis of my 8-channel work *Temple*, created with three individual voices and an IR recorded in the San Bartolomeo Cathedral in Bussetto, Italy, an ornate Baroque space with complex reverberation. The three singing voices were mixed with each other in simple unisons and chords to simulate a choir, and each mixture was processed in two ways, one with just the IR, and the other convolved with itself (i.e. auto-convolved) and then convolved again with the IR. These latter versions doubled the length of the sound, thinned out their spectra by emphasizing the most prominent frequencies, and softened the attacks of each note. Then both versions were mixed together, synchronizing their beginnings. The auto-convolved sounds seemed to be halfway between the original sound and its reverberant version, what I described as a 'ghostly after-image.' The combination of high realism with this abstracted version provided an aurally attractive image for the listener, suggesting something quite unique about the space. By the end of the work, only the auto-convolved components are left, leaving the listener in an aethereal acoustic space that suggests an altered perceptual reality.

6. Extending processing into an imaginary virtual soundscape

The next step in this process for me has been to convolve independent sounds with each other, which I refer to as hybrid convolution. Once again, it started in 2009 as an experiment to explore something I was curious about. I had probably tried an example before, but the typical result of convolving two more or less broadband sounds is, not surprisingly, a thick, undifferentiated broadband texture of little aural interest. However, in this case, a fortunate choice of the materials I chose produced something remarkable, enough to inspire a new 8-channel work, *Chalice Well* (2009) (Truax 2011).

One contributing factor to the success of the experiment was that I re-used water sounds from my piece *Island* (2000), specifically splashes from a well recorded by David Monacchi in Italy which featured a strong resonance. Other water sounds, such as a river, rain, a trickling stream and a domestic faucet were also included. When these were convolved with the well sounds, they took on its spatial qualities, as well as softening the hard edges of even the domestic water stream. Moreover, the percussive drops of the splashes in the well each seemed to trigger a wave of the convolved textures, thereby producing a more continuously evolving sense of flow. I next tried convolving the water sounds with granular synthesis textures (as used in *Riverrun*) and again, the dry synthetic granular material became similarly environmental in nature.

The reason why the results of such convolutions were so aurally convincing seemed to be that all of these sounds had a particulate quality, acting like small impulses similar to how many environmental textures are created. The results continued to be convincing even when I expanded the material to non-watery sounds, mainly percussive material such as breaking glass, bubbles, locks and hard consonants. Once again, this hybridization produced interesting families of textures, and when textured sounds were convolved with others, even more complex textures resulted. Their inherent spatial features were produced because when a ‘wet’ (i.e. resonant) sound was convolved with a drier sound, the result appeared in the middle ground; likewise, wet with wet appeared more distant, and dry with dry remained in the foreground. Combined with 8-channel spatialization of 8 simultaneous tracks of related variants, an entire – albeit imaginary – soundscape was created.

In order to give these materials a larger structure, I thought of wells that I had actually visited, and one, Chalice Well in Glastonbury, stood out, not because there was any sound to be heard, but because of a kind of aura it gave off, no doubt suggested by its rich history of myths and legends from this area in southwest England. One of those myths suggested there were caverns beneath the well – never actually discovered and visited – and that this is where Joseph of Arimathea buried the Holy Grail in order to protect us from the underworld. Sceptics, of course, have pointed out that many of these legends were invented by the monks of the time to promote tourism, an effect that has lasted to this day. Mythical or not, the well provided an appropriate set of imagery on which to base my imaginary water-filled caverns, and to structure the piece as a descent into them (a vertical element that can only be suggested), passing through various caverns, encountering the underworld whose evil is quelled by an aural version of the Grail.

The softening effect of convolution, and the abstractedness provided by hybridization also supported the virtual quality of the Chalice Well scenario in other ways. For instance, another sound source was a short phrase (about ‘a well of flowing water,’ from the Song of Solomon spoken by a female voice) that was also convolved with the water and the other percussive sounds. The vocal formants were extended in this process (the words being unrecognizable) and coloured the environmental sounds, thereby connecting to the traditional gendering of the well as feminine. These sounds were moved around the 8-channel space in circular trajectories, so they seemed to float above the water, and this section is titled 'The Chamber of the Feminine.' Likewise, a section called 'The Glass Chamber' features hybrid convolutions between glass breaking and the other source material. The imaginary quality of these sounds seems to evoke a mythical, even magical quality to what otherwise seems to be a realistic water-filled cavern.

A more recent piece, *Rainforest Rave*n (2020), returned to this hybridization process by convolving dripping water with a windchime, and percussive rain on a roof with a gong. In each case, I used a so-called moving convolution window at a micro level, namely 50 and 100 ms for each stereo channel, respectively, such that the natural sounds were coloured moment by moment by the musical timbres without producing simply a broadband texture. The emotional character of the former is joyful, and that of the latter is very dark and sombre, a trajectory seemingly guided by raven cries that become increasingly processed, and clearly influenced by the current pandemic experience. The piece optimistically ends with the return of a new day and a beautiful natural soundscape.

7. Conclusion

Aside from this brief summary of my personal compositional trajectory with soundscape composition, I have also attempted to trace a path centred in listening that necessitates an intellectual shift away from traditional disciplinary approaches to sound that is based on energy and signal transfers. The complexity of listening in context inevitably leads us both to other areas of study and beyond the signal transfer model, for instance to how sound functions as an ecosystem and acoustic habitat where everything is interrelated. Although humans have always tended to be anthropocentric, as exemplified by visual perception, the aural perspective tends to be more holistic in that we feel embedded in a constant flow of spatial, temporal and social relationships created by sound.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

Barry Truax is a Professor Emeritus in the School of Communication at Simon Fraser University. He worked with the World Soundscape Project, editing its *Handbook for Acoustic Ecology*, and has published a book *Acoustic Communication,* that deals with sound and technology, now in its 2nd edition. In 1991 his work, *Riverrun*, was awarded the Magisterium at the International Competition of Electroacoustic Music in Bourges, France. Truax's multi-channel soundscape compositions are frequently featured in concerts and festivals around the world. In 2015–2016 he was the Edgard Varèse Guest Professor at the Technical University in Berlin. Website: www.sfu.ca/~truax

References

Drever, J. L. 2017. "Field Recording Centered Composition Practices." In *The Routledge Companion to Sounding Art*, edited by M. Cobussen, V. Meelberg, and B. Truax, 253–264. New York: Routledge.

Krause, B. 2012. *The Great Animal Orchestra*. New York: Little, Brown.

Monacchi, D. 2016. "A Philosophy of Eco-acoustics in the Interdisciplinary Project 'Fragments of Extinction'." In *Environmental Sound Artists in Their own Words*, edited by F. Bianchi and V. J. Manzo, 159–168. New York: Oxford.

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

Sterne, J. 2012. *Sound Studies Reader*. New York: Routledge.

Truax, B. 1988. "Real-Time Granular Synthesis with a Digital Signal Processor." *Computer Music Journal* 12 (2): 14–26.

Truax, B. 1999. *Handbook for Acoustic Ecology [CD-ROM]*. Burnaby, BC: Cambridge Street Publishing. www.sfu.ca/~truax/handbook.html

Truax, B. 2001. *Acoustic Communication*. 2nd ed. Norwood, NJ: Ablex.

Truax, B. 2002. "Genres and Techniques of Soundscape Composition as Developed at Simon Fraser University." *Organised Sound* 7 (1): 5–14.

Truax, B. 2008. "Soundscape Composition as Global Music: Electroacoustic Music as Soundscape." *Organised Sound* 13 (2): 103–109.

Truax, B. 2011. "Sound, Listening and Place: The Aesthetic Dilemma." *Organised Sound* 17 (3): 1–9.

Truax, B. 2017. "Acoustic Space, Community, and Virtual Soundscapes." In *The Routledge Companion to Sounding Art*, edited by M. Cobussen, V. Meelberg, and B. Truax, 253–264. New York: Routledge.

Truax, B. 2019. "Acoustic Ecology and the World Soundscape Project." In *Sound, Media, Ecology*, edited by M. Droumeva and R. Jordan, 21–44. London: Palgrave Macmillan.