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

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 it i takes a



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. 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)



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

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 one's ears, or embedded into an artificial head (or u t) 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

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 i ut 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 s34. 5tvnte(t (-75 ()pntp fomrmhav() 0.3 (h)pecntsng.



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

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, i W (2009) (Truax 2011).

One contributing factor to the success of the experiment was that I re-used water sounds from my piece (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 i u) 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, i t 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,

Krause, B. 2012. T t i t . New York: Little, Brown.

Monacchi, D. 2016. "A Philosophy of Eco-acoustics in the Interdisciplinary Project 'Fragments of Extinction'." In i t S u ti t i T i W , edited by F. Bianchi and V. J. Manzo, 159–168. New