Barry Truax

Voices in the Soundscape:
From Cellphones to Soundscape Composition

Abstract

Speech is arguably the most basic form of human communication, but since the 20th century it has become an increasingly disembodied element within the soundscape through amplification, recording and broadcasting. However the same technology that eliminates the face-to-face role of speech within the acoustic community, making it part of the soundscape, has also provided the means for a creative use of vocal material in electroacoustic composition. Just as the concept of "soundscape" embraces all forms of sound and emphasizes how sound is understood by listeners, so too soundscape composition creates simulated environments of sound within which the distinctions between voice, music and environmental sound are blurred. Examples are drawn from the work of the World Soundscape Project at Simon Fraser University and the author's electro-vocal soundscape compositions.

1   Introduction

The human voice—the voice of any species in fact—is arguably central to acoustic communication and our interaction with the world. In this paper I will argue that its physical characteristics in terms of pitch, spectrum, loudness and temporal scale, not to forget psycholinguistic silences (Bruneau 1973), form a reference for all other sounds we encounter in the soundscape, sounds which are higher, lower, louder, softer, vocal-like or not, all of which are judged in reference to the human scale. The complex interaction of the voice, and the other sounds we make with our body and tools, with the immediate environment in the form of timbral colouration, reverberation and so on, provides crucial feedback about our relation to what I call "acoustic space", the sense of space created by all the sounds in a soundscape. Unlike the visual sense of space which appears stable, acoustic space is highly dependent on temporal events and therefore constantly in flux. In other words, the interaction of voice and acoustic space provides us with one of our most basic orientations in the world, and therefore is central to

the human aspects of acoustic ecology. When that interaction is minimized as in an anechoic chamber with its utter lack of reflective surfaces, or when that interaction is overwhelmed in a diffuse sound field with its excessive reverberation, or in a high density noise environment, we may feel disoriented and seek out the means to either restore a sense of balance (e. g. earplugs, portable audio devices) or look to other senses in compensation. On the other hand, when sonic interaction is encouraged within an acoustic environment, we can achieve a balance between listening and soundmaking.

If we shift our attention from the individual's orientation to that of a larger social group, pervasive and complex acoustic interactions create what I call an "acoustic community" (Truax 2001), that is, a bounded space where the shared experience of everyday sound creates identity and dynamic relationships, a system that can be regarded as an acoustic ecology. Soundmaking in pairs and groups of all sizes acts as an important reference for other types of social interaction and often comes to symbolize them. One can understand much of a culture and its social structure by the way it organizes soundmaking within the community. Further, when we examine how communities change and evolve over time (Järviluoma et al. 2009), change will be reflected, and perhaps even predicted, by changes in the soundscape. The impact of those changes on the voice will likely indicate what social, cultural, psychological and political effects are at work.

As complex as such acoustic interactions are, and as inevitable as change is in the soundscape, the impact of electrification in the past century or more has arguably been stronger than any other change in the previous millennium. The inviolable constraints that govern the behaviour of acoustic sound are broken once electrical energy is added. Sound can now be louder, last longer, be stored, transmitted, processed and reproduced arbitrarily. Although the theme of this paper is primarily the impact on soundmaking—the "electrified voice"—the resulting changes are no less profound on the listening process. How to make sense of reproduced and processed voices (as well as music and other sounds) requires the development of new listening skills. Sounds are experienced in a place, but they are not necessarily *of* that place; we can refer to them as displaced and disembodied, or perhaps more accurately, re-embodied via the loudspeaker. Amplified and reproduced voices create a more complicated (but not necessarily more complex) communicational situation because of the contradictions they introduce. Is the source of the voice a real person or an inanimate loudspeaker or is it entirely synthetic; it can be all three at once.

The impact of such electrification on the acoustic community can be profound. At the personal level we see the development of new, more analytical listening skills designed to understand the emerging forms of electroacoustic communication, as well as an increased reliance on what I call "distracted lis-

tening" (Truax 2001) to sound that acts as an accompaniment medium, what is essentially a surrogate aural environment that is possibly of one's own choosing. At the community level, we see dramatic shifts in political and economic power as sound and ways to experience it are commoditized and sold. Digitization adds another level of complexity to the situation as the voice (and all other sounds) is reduced to data that can circulate freely (at least in principle) on the Internet. In the face of such complexity, what models can researchers use to understand these changing forms of acoustic communication?

2   Models of Acoustic and Electroacoustic Communication

2.1   The Energy Transfer Model

The most traditional and pervasive model of how sound communicates is the energy and signal transfer model (Truax 2001). It conceptualizes acoustic energy as travelling (or, more properly, propagating) from a source through a medium to a receiver. It is a highly linear model with a one-way flow of energy. Classical psychoacoustics sought to understand how the "receiver", namely the human auditory system, analyzes the acoustic energy received by the ear to produce spectral and temporal information, mainly in the inner ear. More contemporary approaches in psychoacoustics investigate the cognitive aspects of auditory information processing in the brain, and slowly knowledge about these processes accumulates. However, it does not change the basic paradigm embodied in this model. Early communication theorists such as Harold Lasswell (as quoted in Leiss 1991: 292) summarized the model as "who says what in which channel to whom with what effect". Electroacoustic intervention in the process retains the linearity of the energy transfer model, but interrupts the propagation of acoustic energy by converting it (more properly, transducing it) into an electrical signal, and more recently into a digital data stream, that can be stored, processed, and reproduced via a loudspeaker (Sterne 2003). I describe this process as the "black box" model of signal transfer (fig. 1), a model that most tellingly is concerned with the supposed "fidelity" or faithfulness between the original sound and its reproduction. The myth of fidelity—which figures prominently in advertising by the audio industry—is one that implies conjunction rather than disjunction, extension rather than transformation, realism versus enhancement and virtuality.

It is therefore not surprising that it was the reproduced voice, along with music, that was the first to exemplify fidelity, most iconically with RCA Victor's adoption of the image of a terrier supposedly listening to "His Master's Voice" (HMV) as reproduced by a gramophone of the day. The faithful dog is seen to be
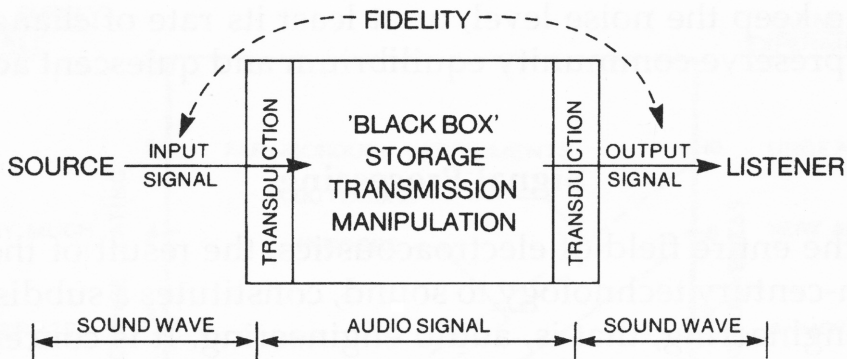
Figure 1: The black box model of audio signal transfer.

recognizing the voice of his master despite what we would now regard as poor fidelity of reproduction. The cocking of the dog's head to imply that it is listening is in fact an anthropomorphic interpretation, as the purpose of that behaviour is actually to create motion in the visual field, not to direct listening. The effect was played out with large human audiences as well with the so-called Edison tone tests (Thompson 2002) from 1915 to 1926 where the reproduced voice of singers could reportedly not be distinguished from that of the original singer who also was present, neither by "experts" or mass audiences to whom this demonstration was presented. Keep in mind that this was the era of mechanical reproduction prior to its electrification (which in fact improved the frequency range and signal-to-noise ratio). I would argue that these experiments were not merely advertising hype or mass hypnosis, but rather evidence that listeners hearing reproduced sound for the first time did not have the analytical listening abilities to detect what now would be obvious differences. They heard a recognizable voice coming from the gramophone horn and that meant it was the "same" as that of the person making it. Later advertising campaigns therefore had to educate listeners to hear—and invest in—the differences that technological improvements in audio reproduction brought about. That is, as the technology changed, so did listeners and their habits of listening, both to the sounds themselves and the social conditions of their reproduction.

At some point the standards for "fidelity" split into that for music (full bandwidth) and that for voice, with "voice quality" audio being of a lower standard, presumably that which minimally allows the voice to be understood. The driving force behind the adoption of this standard appears to be telephony. The frequency range of the inexpensive carbon microphones used in the telephone receiver were found to be adequate for speech to be understood, that range being 300 Hz to 3 kHz, even though the vocal frequency spectrum is about

100 Hz to 10 kHz. This reduced bandwidth allowed more signals to be routed simultaneously though the transmission system, which greatly improved its economic viability. For the listener, speech remained understandable because the main formant frequencies of the vowels were being transmitted, including those in the critical 2 to 3 kHz range, plus sufficient consonantal information, though the higher frequency sibilants (e. g. s, ch) were missing. At the lower end of the spectrum, the psychoacoustic phenomenon of the "missing fundamental" which fills in the perception of the lowest pitch even when it is weak or missing, meant that even male voices were recognizable as such. In addition, the lack of low frequencies below 300 Hz meant that the voice sounded more distant and lacking in room resonances. To listeners, a distant sounding voice corresponded to their knowledge that the person they were talking to was physically distant, and simultaneously preserved the apparent interpersonal distance that would normally be expected (as opposed to having someone speaking right at your ear). Although telephone users initially sometimes shouted to get their messages across the perceived distance, they quickly learned to adapt their vocal style to a more conversational level.

The same process of interaction between the industry and the behaviour of listeners negotiating their uses of technology continues to this day. The last era of claims to increased fidelity was actually the 1980s with the introduction of digital audio in the consumer market with the Compact Disc; in fact, over the past decade, the most pertinent development has been to convince consumers that a *reduced* level of fidelity in the form of compressed audio (typified most strongly in the mp3 format) is the accepted norm. Like telephony in the early days, the compressed mp3 format is justified by the psychoacoustic phenomenon that certain frequency bands need not be transmitted to retain musical quality (Sterne 2006). Interestingly enough, the cellphone service company Fido has relied throughout its advertising history on the look-alike image of dogs and their masters in a long tradition dating back to HMV (i. e. Fido = fidelity). Using a slogan such as "you are the master" Fido attempts to convince cellphone users that they are in charge of their acoustic communicational practices, rather than being tethered to universal accessibility wherever they are.

The competing symbol to the orality-based HMV image throughout the past century is the visual and literate image of the "recording angel" (Eisenberg 1987) that shows an angel or cherub using a quill to inscribe sound into the grooves of a record or the optical pits of a Compact Disc. The "divine" nature of this image implies a perfect inscription, much as the legend of Gregorian chant being dictated to Pope Gregory by an angel implied a perfect form of communication. The visual clarity of images, free of distortion, forms an appropriate representation of a supposedly noiseless audio form of reproduction, mapping the

eternal struggle of good and evil onto the equally eternal struggle between signal and noise in any audio system.

The technical and economic aspects of these marketing campaigns and their effect on listeners and consumers, now being one and the same, masks the profound effects of how these products are culturally used, and ultimately what constitutes the new forms of electroacoustic community. Once again, the voice is a reference point. Murray Schafer (1969) refers to the "schizophonic" split between the original sound and its reproduced form, as well as the context of its reproduction, and to be sure, fidelity is hardly the issue when we think of the larger-than-life amplified voice, often bass boosted, or the smaller-than-life voices we hear reproduced through miniature loudspeakers. Moreover, once a new technology is introduced and the initial phase of social disruption is over, that technology becomes accepted and re-integrated into acceptable social norms (e. g. where and when can one talk on a cellphone or play a portable audio device). The nervous "aberration" implied by Schafer gives way to a banal acceptance, and even the ability to disregard its presence unless needed. Perhaps the most dramatic example of a re-negotiation of acoustic space today comes with the pervasive use of the iPod (Bull 2000, 2006, 2007) as an alternative to being unwillingly exposed to both noise and background music in the soundscape. Listeners now feel they have a viable alternative, the most important aspect of that shift being the sense of personal control it provides. Whether such listeners will also successfully re-negotiate the social isolation it creates with strategies that don't entirely cut them off from others remains to be seen, but compensatory practices seem to be emerging among some users.

2.2   The Soundscape Model and Neo-Orality

As a deliberate contrast to the objective energy transfer model, Schafer introduced an entirely subjective model for soundscape studies that is listener centred (Schafer 1977). Instead of the one-way flow of energy from source to receiver, the most basic soundscape model posits a two-way interactive relationship between listener and environment based on listening and soundmaking. The ideal situation is when those two processes are balanced. This can be as simple as hearing one's own sounds during a soundwalk, or it can describe the complex interactions of an acoustic community.

Since the voice reflects the whole person, both physically and psychologically as well as through cultural conditioning, the feedback we get from hearing our voice is crucial both to our orientation to the given acoustic space and to our sense of identity. In other words, voice and soundmaking mediate one's relationship to self, self-image, gender, and to the environment and society at large,

as well as leading to interactions with others and larger social groups. This feedback can be disrupted by the acoustic situations already described (anechoic, diffuse or high noise conditions), as well as by hearing loss and the use of earplugs or headphones, the effects of which are opposite to each other (i.e. we speak more quietly with earplugs since the voice comes to us largely through bone conduction, and more loudly with headphones because of the acoustic barrier).

Amplification and recording of the voice changes the feedback process as well, particularly for the individual. The amplified voice comes back to the speaker later, usually altered in its timbre and possibly louder than normal. If the delay is substantial (e.g. a quarter second) it can cause the speaker to stutter or even cease speaking, an effect that is particularly strong with headphones. The best strategy is to slow the speed of speech under amplification, particularly in a reverberant environment, but inexperienced speakers are seldom used to doing this. Likewise, vocal self-image can be distorted by hearing a recording of one's own voice since it lacks the head and bodily resonances associated with the normal acoustic process. People are usually embarrassed by the sound of their voice on a recording the first time they hear it, and possibly never get used to the experience. The visual equivalent would be mirrors and photographs that ironically represent the individual in opposite orientations, switching left and right sides, and therefore people often say that photographs don't look like themselves since they are used to their mirror image.

Under such conditions of audio transfer, professional speakers learn to adopt a more formalized type of voice in order to project whatever image is desired. This can be the more authoritative voice of the newscaster, the stylized patter of the deejay, the ritualistic voice of the amplified auctioneer, or the persuasive voice of the preacher, among many other examples. Skilled actors in commercials and dramas can affect a voice to suit the character involved, ranging from the highly cartoonish to one of overwrought emotionality. Listeners are so used to identifying the character and emotional state of a person from their paralanguage (i.e. the non-verbal qualities of speech) that in the visual absence of the speaker, such identification must be done quickly and unambiguously and hence many of the familiar mediated voice types are stereotypical. On commercial radio, for instance, the voice quality of the announcer reflects the demographic of the target audience (i.e. someone one can relate to) and the product image being created in the advertisements. I would also argue that when the form of the verbal communication, mainly its paralanguage, is perceived to match and be appropriate to the content of what is being said, the listener is most likely to find the speaker believable, genuine, sincere and trustworthy. If not, one detects irony, sarcasm, teasing, scepticism, hidden meaning, falsehood, misleading in-

tentions, withholding, or manipulation, but that doesn't necessarily mean one isn't susceptible to their effect.

An interesting way to evaluate the impact of electrification on the voice is to compare Walter Ong's catalogue of the characteristics of primary orality (Ong 1982), not merely with his lucid comparison to print, but to what he termed "neo-orality" or "secondary orality" as created by contemporary media (table 1).

Table 1. A comparison of Walter Ong's model of orality (Ong 1982) with the neo-orality of contemporary media.

Orality	Neo-Orality
Aural (in time, evanescent; interior)	Audio (in time, repeatable, disembodied)
Internal memory, mnemonic (narrative, myth, poetry) – rhyme, rhythm, alliteration, assonance – formulas, cliché, slogan, epithet	Cultural memory (icons, brands, jingles) – all oral mnemonic devices – logo, stereotype, tag line, hook
Wholistic, analogical (riddle, proverb)	Gestalt, symbol, image (ad image, story)
Involved, subjective, "agonistically toned" (empathetic, participatory)	Resonance or avoidance ("relate to", apathetic, cynical, resistance)
Words have power, are events in themselves (naming)	Brand names have iconic power and associations
Cyclic time → repetitive	Cyclic templates & formats
Integrative (groups, communal activity)	Demographic markets (target consumer groups)
Redundant, copious, additive	Sound-byte, repetitive
Conservative, homeostatic, traditional	Turnover of new "improved" products Collapsing of historical time into an "eternal present"

Both the traditional and contemporary forms of orality are designed to promote memorability, though for divergent cultural goals. All of the mnemonic devices of primary orality are heavily used in advertising media and are of particular importance for the distracted listener engaged in a simultaneous task. Although the content of the texts in advertising is important, a purely textual analysis may miss the aural image that is readily apprehended by a listener whose attention is focused elsewhere. The simplified and repeated advertisement, for instance, can become memorable with its associated product image through frequent exposure and the elusive "catchiness" of an ad. Iconic images (both aural and visual), brand names and jingles are designed to stick in the mind, and if successful, become part of the collective cultural memory shared by large numbers of people, usually of a specific age and demographic, and occa-

sionally across subcultural groups. However, the empathetic and participatory effect of primary orality as practiced by the storyteller may meet with the opposite reaction—apathy, cynicism, and resistance—by the media-saturated contemporary listener because of the perceived manipulation involved. It is therefore the task of the advertiser through the ad agency to test the reactions of consumers in focus groups; hence the high degree of sophistication one observes in advertisement design, particularly at the corporate level.

Where primary and secondary orality differ, other than the obvious commercial intentions and practices of the latter, is largely in the structural differences in presentation between the two traditions. Contemporary audio media rely on exact repetition, the position of ads within the highly structured programming flow, and instead of the redundant and copious forms of orality, the compact form of the contemporary sound byte. These structural aspects have been proved to be more effective with audiences that have shorter attention spans, are typically distracted, and are using media as an accompaniment environment, not as their primary focus.

In summary, the soundscape model posits a dynamic interplay of listening and soundmaking, both for the individual and the community. The impact of audio changes the form and uses of electrified sound, and listeners are challenged to make sense of the new forms of communication and to adapt their own uses of that technology in order to achieve greater personal control, particularly in the face of the economic power wielded by the media industries to influence consumers. By exploiting many of the traditional techniques of primary orality, electronic media bypass print and seek to communicate their messages more directly and influence consumer behaviour. However, audiences are not passive and may accept, resist or re-negotiate the intended messages and media uses (Leiss 1991).

The capitalistic economy depends on a constant turnover of new "improved" products and services, and therefore its audio products require a certain impression of uniqueness or at least difference. Yet, to create a mass market, these products must build on what people already know and are familiar with. As a result, the evolution of audio forms of communication to greater sophistication and apparent complexity is constrained by the economic imperative of the need for advertising to create the audience for these goods. Two related questions arise that have not yet been addressed: where do marginalized voices fit in, and what is the role of artistic creation that wishes to provide alternative messages and experiences? We will return to these intriguing issues in the final section of the paper.

2.3   An Information-Based Communication Model

In order to avoid the polar opposition of the objective energy transfer model and the subjective soundscape model, I have proposed an information-based communication model (Truax 2001) that attempts to integrate the insights of both the other models. Neither of those models is inherently wrong; in fact, each is consistent in its approach according to the assumptions on which it is based. Like all models, they have limitations that can be traced back to those initial assumptions. The communicational model, which has its own limitations too, suggests that we understand the listener's relation to sound as one which is based, first, on the information the listener extracts from sonic experience, and secondly, on the relationship that sound creates between the listener and the environment. The former I term the "micro" model (fig. 2) and the latter the "macro" model (fig. 3).
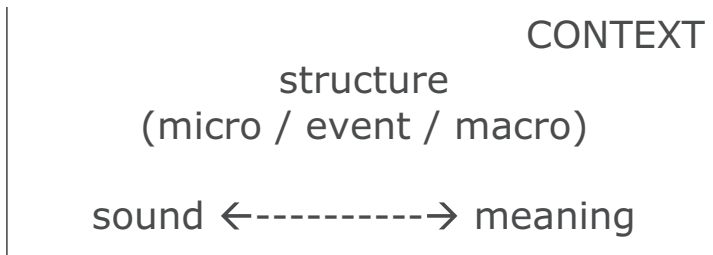

Figure 2: The "micro" model of how meaning is derived from sound.

In the "micro" mode, there are two sources of information available to the listener: that derived from detecting patterns at various temporal levels from the sound itself, and that derived from the listener's knowledge of context. In terms of the voice, the patterns of speech include not only the semantic content of the words, but their paralanguage which often indicates the underlying intent of the words more accurately than the words themselves. They provide not only the emotional "tone" of the communication, but often reflect the implied relationship between the speaker and the listener, whether intimate, friendly, antagonistic, equal or unequal in status and power, and so on. In many cases, the listener's knowledge of context provides the most important clues as to the meaning of speech, and that contextual knowledge can operate at many levels, from psychological to social and political, from personal to cultural. It is often noted how cross-cultural differences can inflect speech and its meaning in vastly different ways, and confuse the outsider who is unaware of its meaning.

The "macro" model is the one that deviates the most from the energy and signal transfer model of speech, in that it suggests that sound does not merely
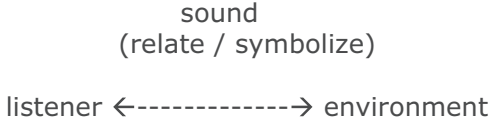
sound
(relate / symbolize)

listener ←------------→ environment

Figure 3: The "macro" model of how sound relates the listener to the environment.

convey meaning, but that it creates relationships. When those relationships become habitual and shared among most listeners in a culture, the sound begins to acquire symbolic importance for the entire culture which in turn affects the reception and interpretation of sound at the micro level. An obvious example of this symbolic relationship is the soundscape concept of the soundmark, a unique and important sound in a particular cultural context that is understood widely, probably due to its history and associations. Interestingly, unlike other signs, a soundmark's actual sonic character becomes inextricably entwined with the symbol, and therefore it is vulnerable to cultural disruption if it is changed through some form of modernization, for instance. One only has to think of the difference between an acoustic bell, an electronic version of the same, and a digital beep to realize the impact of a changing sonic form despite the fact that the sound might mean the same thing. The sign, in the aural case, is not arbitrary as it is in language, but is the result of the interaction of all levels of the communicational models described above.

Another example of this complex interaction of levels comes when we remember, for instance, a particular voice, whether of personal or cultural importance or both. Unless there happens to be some typical phrase or utterance that was repeatedly experienced, we usually remember some kind of "image" conveyed by the voice that brings to mind the relationship we had or have with that person. It matters less what they said, than how they typically said it (i.e. their paralanguage) and how that paralanguage related to the verbal content as well as to the relationship we experienced. With audio media, such voices can become incredibly familiar despite our never having met the person (and if we do, we often experience a disconnection between what we imagined them to be like and what they really are). We think of the "media" as the technical means of information transfer, but it is the voices, the music and other sounds they convey that actually mediate and inform our relationships to what those voices represent, or are attempting to represent (e.g. the image of authority, commodity image, and knowledge in general).

In other words, the audio medium allows those controlling it to create and impose meanings that the listener negotiates, in the sense that they can be accepted, rejected, misunderstood or mistrusted. Acoustic communication at this level creates its own reality, being unhampered by being tied to the real world of acoustic sources. Admittedly rhetoric and storytelling in all its traditional

forms attempted to engage, persuade and entertain the listener, but the creative means available today far surpass what was traditionally available, and in fact, as we have seen with neo-orality, those means utilize all of the traditional techniques and add further complexity through technological manipulation. We now return to the question as to whether these media are designed to reinforce the power of institutions and the goals of the mass market, or whether the audio medium allows alternative voices to emerge and be heard. The technology itself does not determine its uses; how it is organized and used at the personal level does.

3   The Alternative Voice

The acoustic community that traditionally emerged from social and geographic forces, in the audio world becomes a market that is largely determined by economic factors. Harold Innis (1972), in his classic analysis of political economic structures of communication, referred to the flow of resources and commodities between the centre and the margins. Raw materials flowed from the margins to the centre(s) and were processed into commodities that flowed back again. Early forms of mass communication such as the radio could also be understood as creating a similar flow. Today, these elements are largely defined economically; the "centre" is the mainstream market, and the "margins" are subcultural niches. The de-centering role of the Internet has been much discussed as to whether it realigns power between mainstream corporate interests and the multiplicity of individual interests, or whether the former seeks to rationalize and exploit the latter with merely an illusion of individuality. It is interesting to compare some specific practices within the mainstream/margin dichotomy with respect to the voice since vocal practice, as we have already observed, seems to be deeply embedded with cultural significance in each case.

Contemporary feminist studies of the electrified voice in mass media have clearly identified the gender implications that those media have established. Frances Dyson (1994) traces what she calls the "genealogy of the radio voice" as an example of masculine hegemony established in that pervasive medium of the 20th century. It was well documented in the 1920s and 30s that radio listeners did not want to hear the female voice in any kind of authoritative capacity, either in news broadcasts, announcers or commercials. In fact, the only examples of the female voice in radio I am aware of are in the female roles in radio dramas, or in ads and testimonials that are miniature domestic dramas associated with a product women are likely to use. Even as late as the 1990s when I analyzed over a thousand radio ads (Truax 2001: ch. 11), the incidence of the spoken female voice in advertisements in Canada remained at 20 % for the solo voice and when

combined with male speakers. However, female voices were more likely to be heard singing the lyrical message associated with the product. Female voices began to take on announcer roles in the last two decades of the 20th century in North America, but in many cases it seems they were expected to sound somewhat masculine with a lower pitch and huskier timbre.

Likewise with mainstream film, Kaja Silverman (1984, 1988) notes that the female voice has been disembodied in classical film and seldom if ever used in the standard voice-over of the soundtrack, a role reserved for the omniscient sounding male voice. She remarks that "the capacity of the male subject to be cinematically represented in this disembodied form aligns him with transcendence, authoritative knowledge, potency and the law—in short, with the symbolic father" (Silverman 1984: 134). The female voice, in contrast, is normally synchronized with her visual image and as a character she is situated within the action.

The experimental art tradition in the 20th century that focused on the human voice, variously termed text-sound, phonetic art, sound poetry or simply sound art, also tended until recent decades to be male dominated. Dadaists such as Tristan Tzara and Kurt Schwitters liberated the spoken word from its literary forms by dissecting it into phonemes and non-verbal utterances. With the proliferation of recording and audio processing technology in the postwar period of the 1950s, many electroacoustic composers began to incorporate speech and vocal sounds into their work as source material to be processed; likewise, some poets also began to use that technology as an extension of their written and performance practice. A classic example is Luciano Berio's 1958 *Thema (Omaggio a Joyce)* that uses a reading by Cathy Berberian of the first 40 lines of chapter 11 of Joyce's *Ulysses* as source material. Berio describes the intent of the extensive processing of this rich vocal material as creating a continuum between speech and music, between word and sound, and between sound and noise, where the expressive character is derived from their fluid transformations (Chadabe 1997).

Sound artists of all kinds use vocal and non-musical sounds to create an artform free of the perceived hegemony of music, whereas electroacoustic composers tend to insist that *all* sound is potentially musical material (Kahn 1999). One of the main distinctions in their work is the degree to which the sounds they use remain mimetic of the real world or become abstracted through processing (Emmerson 1986). Given the human auditory system's abilities to detect human vocal qualities, even amidst noise and distortion, and even to ascribe those qualities to sounds of non-human origin, it is unclear whether any sound of vocal origin ever becomes entirely abstract. A great deal of processing would be required for that to happen, but it is not impossible.

Cathy Lane (2006) has proposed a useful classification of the uses of voice and

text in electroacoustic music. First is the source of the vocal text, whether scripted and read, derived from interviews or conversations, or retrieved from archival sources. Next she outlines possible compositional techniques used in spoken word work, such as the dissolution or accumulation of meaning using various techniques. Other categories document the role of speech in the work, such as narrative content, narrative guidance or suggestion, phonetic excitation and melodic or rhythmic extraction. Lastly she deals with the semantic quality of the words, such as juxtaposition and permutation, nonsense and new languages, as well as computer-generated speech or computer-composed text. These highly varied practices demonstrate the many levels on which composers and sound artists can deal with the voice.

Gender bias can still be found in electroacoustic music, though it is not as obvious as in the mass media, partly because the human voice can become ambiguous through processing. Hannah Bosma (2003) has surveyed a sizeable repertoire and found that whereas with works on fixed media (e.g. tape or a digital file) there is a balance of gendered voices, in mixed works (i.e. those combining a live performer with an electroacoustic accompaniment), female performers outnumber male performers, just as male composers outnumber female. She attributes this imbalance to the traditional association of composition as a masculine activity involving rationality and innovation, whereas performing is associated with the feminine trait of emotional expression. Adding technology to the mix only reinforces the stereotype of the male composer hidden and in control of the technology while the female performer is on display with her virtuosity and expressivity. Fortunately, the emergence of various composer-performers, as well as feminist, gay and lesbian artists providing alternative forms of expression, is changing this landscape (Truax 2003).

In my own work as a composer, voice and text have been central to a large percentage of my work for over 40 years (Truax 2008b, 2010). This includes many works for singers, ranging from solo voice with tape through to large scale operas, as well as numerous pieces where vocal material is the predominant source. In the mixed works, the tape component is often created with vocal material as well, in order to create a unity between the embodied and disembodied voice, and to use that contrast in a meaningful psychological or even symbolic manner. My development and use of granular time-stretching (Truax 1992, 1994) has provided the means by which the spoken voice can be extended in time from small stretches that turn speech inflection into something closer to singing through to large stretches that turn it into something that suggests an environmental sound, once its duration exceeds that of the human breath. Through multi-channel reproduction, such vocal sounds can create an immersive environment for the listener who seems to be placed *within* the voice, as it were.

A relatively simple example can be found in *Song of Songs* (1992), for oboe d'amore, English horn and stereo tape with optional computer graphics. The sensual text from the Song of Solomon, recorded by both a male and female speaker, is stretched into a song-like version (whose melodies are imitated by the live instruments), just as the sounds of birds, cicadas and crickets are slowed down to the point where their temporal patterns are perceived as being closer to the human scale. In other words, all humanity and nature appear to be singing, and in fact the piece also includes a recording of a monk singing, and quotations in the instrumental part of the Hebrew cantillation melody associated with the text. This strategy reflects the extended metaphor of the text comparing the beauties of nature to those of the beloved, and vice versa. By keeping the gendered references in the text intact when spoken by both male and female voices, the piece becomes inclusive of both heterosexual and homosexual desire, just as the Christian and Jewish traditions are merged. The theme of love blurs all such arbitrary boundaries.

Another approach that I have extensively used with digital granulation techniques is to isolate small vocal fragments on the order of the phoneme and to extend, process and mix those until they create a large-scale texture. In fact, the first work to do that, *The Wings of Nike* (1987), used only two phonemes (about a third of a second of material in total), one male and one female, to create three movements of about 12 minutes duration in which one hears everything from isolated fragments to rhythmic figures, densely pitched textures and finally the obliteration of pitch into broadband noise when the duration of the unit grain is shortened to a few milliseconds. The visual counterpart to the soundtrack is a series of digitally processed images by Theo Goldberg derived from the statue known as the Winged Victory or Nike, housed in the Louvre. The larger-than-life quality of the soundtrack (though remaining identifiable as a fusion of gendered voices) is designed to complement the heroic figure represented by the statue.

Digital resonators can be used to model various instruments, the simplest being a one-dimensional delay line that models either a string fixed at each end or a tube open at both ends, or else closed at one end. My pair of pieces, *Wings of Fire* (1996) for cello and tape, and *Androgyne, Mon Amour* (1997) for double bass and tape, are based on poetry by Joy Kirstin and Tennessee Williams respectively. The spoken poems, and processed versions of that material, are heard on tape, often processed through resonators that are models of the live instruments, and tuned similarly. The effect is to bridge the gap between the voice and the instrument, with the signification that the love poem is being addressed to the instrument as the performer's lover. A theatrical element, particularly in the latter piece, adds to the narrative element implied by the poetry.

Finally, another pair of pieces, *Temple* (2002) and *The Shaman Ascending* (2005–2006), are exclusively based on sung vocal material. In the former, the

voices of a counter-tenor, alto and bass are processed with the technique of convolution in two different ways. In the first, often called impulse reverb, one convolves the voice with the impulse response (e.g. the breaking of a balloon) recorded in an actual space, in this case an Italian cathedral in the town of Busetto. The result is that the voice appears to be in that space. When distributed over 8 channels surrounding the listener, the effect is strikingly similar to being in that reverberant cathedral. The second use of convolution is to convolve the sound with itself (i.e. auto-convolution) which results in a doubling of its duration and a thinning of the timbre to emphasize the strongest frequencies and de-emphasize the weaker ones. The result is a sound that hovers between the source, the human singer, and the reverberation itself, an effect I refer to as a "ghostly after-image". The two types of processing are heard synchronized, such that one seems to hear a realistic choir (based on a mix of the source voices) where after each note or chord the after-image prolongs the reverberant texture. Finally at the end of the piece, only ephemeral textures of vocal harmonics are left as an indicator of spiritual transcendence.

The sung bass voice in *Temple* is used as the sole source for *The Shaman Ascending* that was inspired by Inuit throat singing and an Inuit pair of sculptures that provided the title of the piece. The sung tones are chopped into short grains, somewhat like *The Wings of Nike* from two decades earlier, but in this case, the grains spin around the 8-channel space in increasing densities, beginning with the impression of fast shamanistic chanting and progressing to complex dynamic textures. Auto-convolution of the material brings out the inner harmonics of the sung material, particularly from the multiphonic style of singing of one of the sources. The purity of these harmonics, as with those heard in *Temple*, signifies the "ascending" character toward an altered state of consciousness that the shaman aspires to. Paradoxically, this reference to traditional oral culture is made possible with sophisticated digital processing from our own era, and the result transcends what is acoustically possible.

These electro-vocal works, as summarized here, can be understood as a subset of soundscape composition, that is, electroacoustic works that are "about" real-world contexts, often using environmental sounds as source materials, but retaining their reference to place and space. Such works can range from the highly mimetic, or what is often called "phonographic" (Drever 2002) because little or no manipulation of the sound is involved, through to the highly abstracted (but not abstract). They combine sonic realism with processed elements that attempt to bring out the inner meaning or symbolism of a specific real-world context (Truax 2002, 2008a). Multi-channel reproduction greatly aids in creating the sense of involvement with a virtual environment. The aim of a soundscape composition is often to focus the listener's awareness to environmental sound and the associated social context it reflects, as well as to create

entirely imaginary or idealized soundscapes that are an aesthetic alternative to our often noisy everyday experience. Voice and vocal material are powerful tools for consolidating those references.

4   Conclusion

In this brief survey of the impact of technology on the voice, we have seen that two contrasting results emerge: the mass market commodity that can be termed music-as-environment where sound and music create an accompaniment environment, one that is increasingly portable and of the listener's choice; and a marginalized experimental art tradition that treats environment-as-music in the form of the soundscape composition. Within those practices, just as with traditional acoustic communication, the voice plays a central and referential role. We can also contrast voice-as-environment, referring to the ubiquity of disembodied voices in our daily soundscapes, with environment-as-voice, referring to new forms of sound art that immerse the listener within a vocal soundscape. The evolution of the traditional acoustic community where the human voice and human activities set the scale of the ecological relationships involved, towards the electroacoustic community where the human voice can be amplified, broadcast, reproduced or synthesized, introduces a new complexity that challenges the viability of our communicational patterns and asserts power and authority as the chief determinants. However, the experience of artistic alternatives, along with individuals negotiating their own relationships to technology, gives us hope that this most human dimension of communication will prevail, no matter how it evolves.

References

Bosma, H. (2003). Bodies of Evidence. Singing Cyborgs and other Gender Issues in Electrovocal Music. *Organised Sound* 8 (1): 5 – 17.

Bruneau, T. (1973). Communicative Silences. Forms and Functions. *Journal of Communication* 23: 17 – 46.

Bull, M. (2000). *Sounding Out the City. Personal Stereos and the Management of Everyday Life.* Oxford: Berg.

Bull, M. (2006). Investigating the Culture of Mobile Listening. From Walkman to iPod. In: *Consuming Music Together. Social and Collaborative Aspects of Music Consumption Technologies*, edited by K. O'Hara & B. Brown. Dordrecht: Springer, pp. 131 – 149.

Bull, M. (2007). *Sound Moves. iPod Culture and Urban Experience.* London: Routledge.

Chadabe, J. (1997). *Electric Sound.* Upper Saddle River: Prentice-Hall.

Drever, J. L. (2002). Soundscape Composition. The Convergence of Ethnography and Acousmatic Music. *Organised Sound* 7 (1): 21–27.

Dyson, F. (1994). The Genealogy of the Radio Voice. In: *Radio Rethink. Art and Sound Transmission*, edited by D. Augaitis & D. Lander. Banff: Walter Phillips Gallery, pp. 167–186.

Eisenberg, E. (1987). *The Recording Angel. Explorations in Phonography.* New York: McGraw-Hill.

Emmerson, S. (1986). The Relation of Language to Materials. In: *The Language of Electroacoustic Music*, edited by S. Emmerson. London: Macmillan, pp. 17–39.

Innis, H. A. (1972). *Empire and Communications.* Toronto: University of Toronto Press.

Järvilouma, H., M. Kytö, B. Truax, H. Uimonen & N. Vikman (eds.). (2009). *Acoustic Environments in Change and Five Village Soundscapes.* Tampere: TAMK University of Applied Sciences.

Kahn, D. (1999). *Noise, Water, Meat. A History of Sound in the Arts.* Cambridge: MIT Press.

Lane, C. (2006). Voices from the Past. Compositional Approaches to Using Recorded Speech. *Organised Sound* 11 (1): 3–11.

Leiss, W. (1991). On the Vitality of our Discipline. New Applications of Communication Theory. *Canadian Journal of Communication* 16 (2): 291–305.

Ong, W. (1982). *Orality and Literacy.* London: Methuen.

Schafer, R. M. (1969). *The New Soundscape.* Vienna: Universal Edition.

Schafer, R. M. (1977). *The Tuning of the World.* New York: Knopf. Reprinted in 1994 as *Our Sonic Environment and the Soundscape: The Tuning of the World.* Rochester: Destiny Books.

Silverman, K. (1984). Dis-Embodying the Female Voice. In: *Re-Vision. Essays in Feminist Film Criticism*, edited by M. A. Doane, P. Mellenkamp & L. Williams Los Angeles: University Publications of America, pp. 131–149.

Silverman, K. (1988). *The Acoustic Mirror. The Female Voice in Psychoanalysis and Cinema.* Bloomington: Indiana University Press.

Sterne, J. (2003). *The Audible Past. Cultural Origins of Sound Reproduction.* Durham: Duke University Press.

Sterne, J. (2006). The mp3 as Cultural Artifact. *New Media and Society* 8 (5): 825–842.

Thompson, E. (2002). *The Soundscape of Modernity.* Cambridge: MIT Press.

Truax, B. (1992). Composing with Time-Shifted Environmental Sound. *Leonardo Music Journal* 2 (1): 37–40.

Truax, B. (1994). Discovering Inner Complexity. Time-Shifting and Transposition with a Real-Time Granulation Technique. *Computer Music Journal* 18 (2): 38–48 (sound sheet examples in 18 (1)).

Truax, B. (2001). *Acoustic Communication* (2nd ed.). Westport: Ablex Publishing.

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

Truax, B. (2003). Homoeroticism and Electroacoustic Music. Absence and Personal Voice. *Organised Sound* 8 (1): 117–124.

Truax, B. (2008a). Soundscape Composition as Global Music. Electroacoustic Music as Soundscape. *Organised Sound* 13 (2): 103–10.

Truax, B. (2008b). A Voice for all Seasons. In: *Playing with Words*, edited by C. Lane. CRISAP/RGAP. Reprinted at www.sfu.ca/~truax/voice.html

Truax, B. (2010). Text and Voice Based Composition. Cambridge Street Publishing, CSR-DVD 1001. www.sfu.ca/~truax/textdvd.html