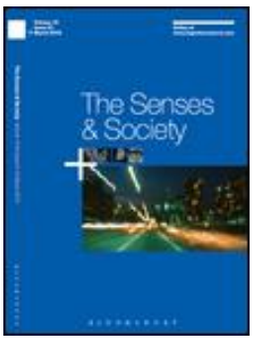
2010

To Know Beyond Listening: Monitoring Digital Music

Ian Reyes
University of Rhode Island, ianreyes@uri.edu

Citation/Publisher Attribution
Reyes, Ian. 2010. "To Know Beyond Listening: Monitoring Digital Music." *The Senses and Society*, vol. 5, no. 2, 2010, pp. 322-338. http://dx.doi.org/10.2752/174589210X12753842356043
Available at: http://dx.doi.org/10.2752/174589210X12753842356043



The Senses and Society

ISSN: 1745-8927 (Print) 1745-8935 (Online) Journal homepage: http://www.tandfonline.com/loi/rfss20

To Know Beyond Listening

Ian Reyes

To cite this article: Ian Reyes (2010) To Know Beyond Listening, The Senses and Society, 5:3, 322-338, DOI: 10.2752/174589210X12753842356043

To link to this article: https://doi.org/10.2752/174589210X12753842356043

Senses & Society VOLUME 5, ISSUE 3 PP 322–338 REPRINTS AVAILABLE DIRECTLY FROM THE PUBLISHERS PHOTOCOPYING PERMITTED BY LICENSE ONLY

To Know Beyond Listening
Monitoring Digital Music

Ian Reyes

ABSTRACT In music production, "monitoring" refers traditionally to audile strategies intended to reveal the "true" sound of mediated audio. Here, it is expanded to include new, digital technologies intended to better know and control the record-object beyond what listening and listening technologies allow. Surveying traditional, contemporary, and emerging tools of record production and distribution, this essay addresses three types of monitoring: audio, visual, and data.

In sum, monitoring entails the supplementation and subversion of the ear through protocols promising to surmount the biases and distortions of audio media. Key technologies include reference speakers, room correction systems, digital audio workstations, open mixes, pre-sets, social networking sites, and automatic music information retrieval. Situating these within a "techoustemology" of monitoring,

Ian Reyes is an Assistant Professor of Communication Studies at The University of Rhode Island's Harrington School of Communication and Media. His interdisciplinary research concerns the aesthetic conventions and social construction of media technologies. Currently, he is researching modes of audio/visual play in video games and other interactive entertainment. He is the author of "The Raw and the Produced: (Re) Creating Hardcore Authenticity in the Studio." ianreyes@mail.uri.edu.

Senses & Society DOI: 10.2752/174589210X12753842356043

the central argument is that many innovations in digital audio are non-auditory and, therefore, displace sound and listening as the central means of producing relevant knowledge about music mediated in the digital age.

KEYWORDS: sound, listening, music, digital audio, monitor, multi-modal perception.

The term "techoustemology" is a hybrid of technology, acoustics, and epistemology. It was coined by Thomas Porcello:

> to foreground the implication of forms of technological mediation on individuals' knowledge and interpretations of, sensations in, and consequent actions upon their acoustic environments as grounded in the specific times and places of the production and reception of sound (2005: 270).

Following a "techoustemological" approach, this essay lays the ground for thinking further about the changing sensorium of recorded music in the digital age. The purpose is to understand efforts to know sound as an object of mass-mediation and to consider how such efforts displace skilled listening in favor of new technologies that are non-auditory by design.[1] This claim rests on an analysis of "monitoring." To monitor is to produce relevant knowledge about mediated audio. Beginning with a description of *audio* monitoring protocols, this essay builds from those protocols to explain the stakes for *visual* and *data* monitoring.

A letter to the editors of *Tape Op* synopsized this matter within an argument for the superiority of recording with digital audio workstations (DAWs) compared to analog technologies:

> It has become second nature for me to look at a computer screen and understand the workings of a song structure as I see it inside of my DAW (Pro Tools LE). The way that waveforms line up and loops interact and beats drop are all visual cues that were never available to the analog engineer. Before, it was tape with EQ and VU meters and listening to all the tracks playing without having a clue about the visual aspects of what was about to come next. These are functions we take for granted in this digital world. […] The concept of mixing and mastering also changes with this visual dynamic. I can mix and master in the box because I know that the relation of my waveforms is directly proportional to its inherent final product as a digital CD or MP3. A blown out waveform is obvious without the need to even hear it. A low gain file is plain as day

> – no play required. I could probably make an entire song out of waveform samples, without even listening to it and know the file will at least, sonically, be proper. (Karalian 2009: 14)

Whether this makes digital superior is beside the point here; the idea is simply that the sensorium of digital music recording is different from the sensoria of earlier modes. Nonetheless, despite the deep transformation of audio media effected by the digital turn,[2] one fact remains: recorded music is *made for speakers*. This informs virtually all aspects of audio mediation, orienting recording toward the final moment of speaker-listening. Yet it takes a certain comportment to the object of speaker-music to create the context wherein non-auditory technologies like visual and data monitoring are desirable.

Consider that there are at least two approaches to making speaker-music, the live-protocol and the record-protocol.[3] Noting that the tools of the trade are more or less the same for producers of live and recorded speaker-music, one sees that the key distinction is found in the way their protocols conceive the relationship between listeners and the aural object of speaker-music. Consider the slight, yet crucial, difference between explanations from Peter Manning (1994) and Albin Zak (2001).

Manning described the problems encountered in the exhibition of electronic art music:

> the diffusion of sounds in large acoustic spaces leads to significant alterations in the perceived images. Many a composer, having realized a work in the confines of a studio, has discovered too late that large-scale projection removes subtleties of texture and dynamics, while spatial effects often assume dramatic proportions. (1994: 359)

Speaker-music of any sort requires monitoring because the life of the work is inseparable from its exhibition. Further, finding that acoustic reality is so inhospitable to *singular* exhibitions of speaker-music, the goal of *mass-producing* speaker-music seems wholly beyond reach. And this would be the case if one insisted on the live-protocol, emphasizing the sounds from the speakers as they operate within the space of exhibition. But the record-protocol reveals another possibility. Zak described the difficulty of creating recorded speaker-music in nearly identical terms, with one key twist:

> The experience of creating a mix that sounds great in the studio and terrible on all other systems is a common one for inexperienced recordists. *A successful mix is one that travels well, rather than one that is perfectly tailored to a particular listening environment*. […] Recordists must continually face the challenge of making *a very specifically configured work with the criterion that it have a general effectiveness in myriad circumstances*. (2001: 117, emphasis added)

Unlike live protocols developed for working in a knowable environment, record protocols bracket-off the recording from the less certain world of all possible environments. So, music doesn't *go through* the speakers, it is *addressed to* the speakers. It's not music for *these* speakers in *this* room, it's music for *all* speakers in no specific place. But for a recording to "travel well," it must also resonate with listeners. It's more than signal transmission; it's about transmitting signals that become significant sounds, sounds that are "sociosentimental."

Theorizing the measure of success for popular music, Antoine Hennion concluded, the "fundamental task [of recording] resides in *the permanent and organized quest for what holds meaning for the public*" (1990: 185). He described the domain of public meaning as "sociosentimental." In music, sociosentimental objects include

> key phrases, sounds, images, attitudes, gestures, and signs, infralinguistic categories which are all the more difficult to pin down insofar as they escape definition by the official language, and are not autonomous but inseparable from the social context within which a given group attributes significance. (Hennion 1990: 186)

One experiences sociosentimentally affective sound when, for example, the genre, artist, or song is identifiable from just a second or two of the recording. Although an analysis of sociosentimental categories is beyond the scope of this essay, the problem of communicating with an unknown public in a poorly understood yet nonetheless highly effective "language" is essential to note as it motivates monitoring protocols. The trouble is that sociosentimental knowledge, though remarkably accurate, is quite hard to articulate and harder still to measure and abstract.

The purpose of monitoring is to know what an audience will hear and which aspects of a recording they listen for, which are ideally but not necessarily one and the same. A "good" recording aligns a material object with a social object. This means there are two domains of knowledge: technoscientific and sociosentimental. This is why the ears of professional listeners, like record engineers, need supplementation. Because audio monitoring is better suited for technoscientific analysis, visual and data monitoring emerge as bridges to sociosentimentality. The horns of the dilemma, using Zak's terms, come from the need to create a "specifically configured work" while guaranteeing a "general effectiveness" in "myriad circumstances." Though this is related to the commercial aims of the music industry, it is necessary but not sufficient cause for commercial success. The point, at any rate, is that a good record conveys its intended aesthetic impact, its sociosentimental message, in all playback scenarios. Accomplishing this, under the record-protocol, requires attention to an object beyond what speakers reveal to listening. Like Jonathan Crary's account of visual

observation protocols "relocating vision to a plane severed from the human observer" (1992: 1), protocols for assisted audile observation similarly lead to the severing of listeners from sound and listening.

Audio Monitoring: Reference Speakers

Audio monitoring is technologically assisted, scientific listening geared toward minimizing bias and distortion. The aim is not simply to achieve a clearer, more complete perspective on the record-object; the purpose of establishing such a perspective is to know the technoscientific components of the sociosentimental system of recorded music. This protocol is complicated because the final recording is for mass distribution and, therefore, beset by an always and already imperfect environment.

Repetition is routinely touted as a fundamental quality of recordings, but for someone monitoring the object to be repeated, the more intense the focus on the object as such, the less repeatable it seems to be. Attention to repetition, particularly variations in the (re)manifestations of a record-object on different playback systems, leads to increased concern for the truth behind sensuous reality. Recording well, technically, is largely a matter of knowing what the speakers reveal/conceal, and assuring that all speakers manifest the most essential qualities of a work.

A recording should still sound right, if not the same, whether it is played from CD, MP3, or vinyl, whether it is heard on radio, through headphones, at home, in the car, etc. To encode for all conceivable scenarios requires skilled listening to and manipulation through audile observation systems like audio "reference" monitors. Reference monitors are specially designed speakers, like microscopes for audile endeavors, used to reveal the smallest details of the record-object. Recordists turn to these in order to establish a zero-point for referencing what a recording "really" sounds like. Monitoring speakers are unlike the speakers on home stereos, for example, in that the latter are made to sound *good* while the former are made to sound *true*. Of course, this is a bit of an oversimplification. The situation may just as well be explained as, consumer audio devices sound bad because they impart more *distortion* than reference monitors, which sound good because they are *transparent*.

Advising novice engineers, Paul Lehrman explained basic audio monitoring protocols:

> What we're hearing through the studio monitors isn't the same as what we hear when there's a direct acoustic path from the sound source to our ears. Ideally, speakers would be totally flat with no distortion or phase error and with perfect dispersion, but even the best monitors are still far from being totally "transparent." (2005: para. 7)

Though all speakers are biased, studio monitors are "truer" because they offer a wider and flatter frequency response than home stereos. What matters is fidelity, neutrality, or transparency as opposed to bias or distortion between listener and object. However, no listening apparatus can be entirely neutral, so it falls upon the ears of engineers to know their instruments well enough to manage whatever bias remains. This means even "neutral" or "transparent" tools for listening are so not by design but by the finely tuned ears of expert listeners.

For example, Yamaha's model NS-10 reference monitors have been a mainstay of the recording industry. They are loved and hated for their clear, sharp sound. The well-known "tissue-paper mod" (i.e. covering the tweeters with a sheet of tissue paper to make them less piercing) highlights the bias of these speakers as well as the fact that some engineers modify, or distort, their instruments to make them less biased or, more accurately, biased in a way more suitable, or useful, to the listener. As useful as they are, however, professional engineers tend to employ *multiple* monitoring systems, as each reveals/conceals different aspects of the record-object.

Bobby Owsinski advised:

> The number of monitor references that are used is an important aspect to getting a mix right. Although a mixer may do most of his work on a single system, it's common to check the mix on at least two (maybe more) other sources as well … the average of all of these systems should make for a good mix. (2006: 67–8)

Multiple monitoring systems work together to produce knowledge about a recording, which the engineer manipulates – mixes – until arriving at one version that sounds right on each of the reference monitors. Although manifestations of the finished work will inevitably vary, multiple-monitoring helps assure variances of reproduction won't compromise the intended sound. Like Stuart Hall's theory of televisual communications, this is an example of how moments of circulation and reception are "incorporated, via a number of skewed and structured 'feedbacks', into the production process itself" (1999: 509). Record-objects that travel well anticipate the plurality of decoding, the relative autonomy of the object, and control for it, encoding a "preferred meaning" by anticipating common distortions of circulation and reception. Because creating a custom object for every environment is impossible, acoustic reality is understood to be a constant problem.

Audio monitors and techniques of studio listening cooperate to solve this problem and isolate the work, focusing solely on those features mattering most for traveling well. Not only are reference speakers specialized for recording with broader and flatter frequency response, most are also tailored for "near-field" listening, meaning

they do not project sound, like a home stereo or public-address system. Instead, they produce a clear stereo "image" just in front of the speakers. With proper placement of speakers and listener, it is possible to reduce room tone (i.e. the effect of sound propagating through space), a major contributor to speaker-music distortion. Again, this means the true sound of a record, under this protocol, does not include the room it is heard in.

As Lehrman explained:

> [E]very indoor space that's not an anechoic chamber has its peculiar colorations, which are different from any other space. We need to be able to compensate for these distortions, consciously or unconsciously, and block out the sound of the speakers and the room as we listen. Our experience and training as professionals teach us how to eliminate the medium and concentrate on the source. (2005: para. 7)

While all sound waves are fundamentally dependent upon space in order to exist *as* sound, the record-protocol makes it possible for the object of recording, the "source," to be separated from the conditions of its mediated observation. While it cannot be separated from speakers without also being separated from sound and listening entirely, the record-object can be isolated to the surface of the speakers absent the room through "experience and training" as a professional listener. Here, knowledge is produced by more than speakers and their positioning; it comes also from the audile skill and self-awareness of the engineer. Good engineers know their ears and speakers intimately.

Yet with new "room correction" systems, speakers can know themselves. These digitally self-tuning reference speakers take a "snapshot" of a room's acoustics and restructure their output to pre-correct for distorting room tone. So, if the analytics find that the audio in the room contains more bass frequencies than the original, pre-speaker digital audio files, then the speakers will reduce their output of those bass frequencies, relieving listeners of the need to listen knowingly. Room correction represents a de-skilling of audile labor in music production, a theme that will be returned to. Whereas room tone was once controlled by an engineer's ability to modify the listening environment with acoustic treatments, position the speakers properly, sit correctly in front of those speakers, then consciously filter out whatever distortions remained in order to know the record-object beyond what's heard, room correction technology renders this object, which had been largely in the mind's ear, as an actual, empirical object for listening.

While self-reflexive audio monitors are impressive, in principle, it is still worth pausing to consider whether this, or any, of today's monitoring tools produce accurate knowledge about the audience's actual listening environment. Lehrman rightly asked

whether engineers ought to be reconsidering their traditional tools and techniques of monitoring due to changes in the audience's environment. For example, today's audience for digital audio listens typically on small speakers encased in plastic as opposed to big speakers in wood cases as in decades prior or monophonic metal horns in the previous era. As the routes change, one has to ask whether the standards for traveling well have also changed. Fortunately, as will later be discussed, knowledge of the listening audience is becoming easier to come by in the digital age. Nonetheless, the dominant protocol remains to isolate the object and to listen in a way that "eliminates the medium." Further, because human listening is a medium in its own right, and therefore prone to biases and distortions, recordists look to supplement listening with isolative means that displace listening from the recording process.

Visual Monitoring: Waveforms, Open Mixes, and Presets

In principle, anyone can make a recording travel well, so long as one has good ears and a known referencing system. In the absence of a wholly unbiased monitoring system, the next best thing is a *known* bias. This problem leads to technologies and techniques designed to know beyond what listening uncovers and, therefore, today, the landscape of recording studios is noticeably different from the past. Computer monitors now sit between reference monitors as another, visual, means of knowing the record-object.

The first playback of "Au Clair de la Lune," one of the oldest known music recordings, in 2008, was a reminder that the earliest attempts at sound recording were wholly oriented towards producing visual records of sound with no expectation of using the image to reproduce the original signal, rather only to know the sound beyond listening. The creator of "Au Clair," Leon Scot, did not anticipate in 1860 that this song would be heard because there were no playback devices; this visual record of sound was intended only as that. Like Thomas Edison's 1874 ear-phonautograph, the point, as the name implies, was only to make sound visually knowable by writing itself. The rise of visual monitoring, then, is a continuation, not a break, with earlier efforts to know sound more objectively by rendering it in a visual format. To be sure, Scot and Edison were not the first to turn sound into images; compositional notation of music is an important precursor to these more direct means for visually rendering sound. One should also recognize that the analog era of recording was not without visual monitors, like volume meters, for example. The difference with digital audio technologies, however, is that the visuals themselves can be manipulated to change the sound. Now, visual monitoring is more than an adjunct to recording, it is arguably the focal point of digital music creation.

As Steve Jones observed, digital technologies brought about a "sensory shift" in audio recording: "[c]omputer-aided music making

moves music into a visual realm, especially during use of digital sampling and recording, and sequencing" (1992: 205). Computer-based DAWs visualize sound waves, drawing their fluctuating amplitude over time like a digital phonautograph; the difference is that these images are more than records of the sounds, they are also interfaces. With the graphical access offered by DAWs, much work can be done without listening at all, including noise reduction, envelope shaping, rhythmic adjustments, and (re)arranging a song's structure. Potentially, visual analysis and graphic editing can go a long way toward ameliorating some speaker-issues on the side of production, empowering amateurs to *look* like professionals even if they lack the proper audio monitors or audile expertise. Further, visual tools aid more than technoscientific analyses, some promise non-auditory guides to sociosentimentality.

Through a DAW, the production of sociosentimental sounds can be studied from the inside out. The Waves Corporation, for example, offers a program called OpenMix, providing DAW users graphically interfaced, technoscientific guides to sociosentimental sounds. OpenMix is not so much software, however, as it is *settings* for software providing access to every element of professional mixes. The package provides complete session files for DAWs, including audio sources as well as, most importantly, the exact settings used to process and render those sources. Because technoscientific particularity of a record-object is the root of sociosentimentality, this is a valuable source of information. Just the fact that this is possible speaks to an enormous change effected by digital technologies. In the analog era, there was no way to save a mix in this suspended state of production. Though some analog-age engineers used visual tools, like diagramming or photographing, to assist in reconstructing a mix, these did not become significant sources of general knowledge. The distribution of open mixes means a user can learn by actually *seeing* the technoscientific components of a sociosentimentally affective song.

A related, commoner, and smaller scale version of this practice, pre-dating open mixes, is the "preset": a previously set configuration of a tool's variable parameters. Digital signal processors (DSPs), a central component of DAWs, are largely organized around presets, permitting easy access to sociosentimental sounds. Further, they can store thousands of settings and permit users to add their own. Though the problem of best matching source to setting remains, the capacity for users to click through thousands of presets – with suggestive labels such as "70s' Guitar Lead," "British Stack," or "Death Metal Crunch" – without needing to know exactly how those settings work enables hunt-and-peck music production. Open mixes and presets lower barriers to entry by de-skilling the labor necessary to produce traveling speaker-music.[4]

The visual references offered by digital audio software raise the question: how should the digital record-object *look*? Though

Lehrman and Karalian suggest relying on visual supplements alone will fail the ultimate test of listening, it's still worth asking how far they can go. Is it possible to *see* whether a recording will travel well? In some regards, yes.

Jonathan Sterne (2006) used waveform analysis to explain the "loudness wars." This controversy among audio professionals has to do with dynamic compression, or raising the average volume of a recording, which makes the music seem better when briefly compared to a track of lower average volume, but, because compression reduces fluctuations of amplitude, such music loses dynamic expressiveness. Sterne's monitoring of waveforms showed the historical progression toward louder recordings. A more typical application of this technology is to monitor dynamics without listening during the recording process, as described by Karalian and Jones. On this, it can't pass without noting that the analog era also had visual volume meters. But, again, the difference with digital forms of volume monitoring is that the visuals themselves are recorded and become a graphic means of affecting the record-object.

Still, well-groomed waves have no necessary correlation to sociosentimental sounds, and presets or open mixes similarly fall short. There's no guarantee, for instance, that the "Death Metal Crunch" preset will actually produce the specific timbre identified with the genre. However, there may be a way to check. Harris Berger and Cornelia Fales (2005) used spectrographic monitoring – visually displaying frequency distribution and amplitude changes over time – to show that the ordering of metal genres along an axis of "heaviness," a sociosentimental quality, is consistent across listeners and linked to audible and visible dimensions of timbre. Berger and Fales' use of spectrographic monitoring proved, in effect, that the rhetoric of metal fans is empirically grounded, that it constitutes relevant knowledge about the music *as sound* and, therefore, is not merely exclusive subcultural slang or empty marketing buzzwords. This means, to know whether a "Death Metal Crunch" preset has produced a verifiably death metal timbre, one could visually monitor the results through comparative spectrographic analysis. Such insights, if they could be uncovered for other sociosentimental categories or genre traits, would not only verify the empirical basis for talk-about-sound but also establish a lexicon of visually generated, spectrographic knowledge about key elements within the language of recorded music.

Still, as productive as these techniques are, in light of the preceding discussion of audio monitoring, one should also ask how, or whether, issues of neutrality play out with regard to visual monitoring. Waveform displays and spectrographs are prone to visual distortions based on image resolution, for example. Further, neither offers much information regarding the source; a loud accordion and a distorted guitar may be indistinct as waves or spectrographs. The best check on visual distortion, naturally, is careful listening. Though listening

is supplemented and displaced, it is still crucial for the affective application of visual tools.

It's unlikely that there will be a wholesale de-skilling of record engineering. However, for the theme of monitoring, this goes to show that protocols of listening are now displaced by visual references and subverted by the potential for plug-and-play, hunt-and-peck music-making taking place almost entirely "in the box" (i.e. within the computer). What's more, the box may become populated with intelligent agents, both human and artificial, capable of monitoring the data constituting digital music files as such. That is, the hallmark of the digital turn is not the visualization of sound, which was already underway prior to the invention of sound recording. What makes digital audio truly digital is the capacity to render sound and listening as data and, further, to inspire new tools and protocols for monitoring that data as a repository of potential knowledge about mediated music.

Data Monitoring: Music Information Retrieval and Networked Listening

Given the lack of sociosentimental knowledge left by the persistent blind spots of audio and visual monitoring, there are two intriguing developments representing nascent protocols of data monitoring: music information retrieval (MIR) and networked listening. Following a trajectory by which *sound* has been relegated to a plane beyond human observation, one finds here the logical continuation of this in the relegation of *listening* to a plane beyond the listener.

Music information retrieval is an emerging protocol intended to handle music databases and the travel of music files over digital networks. But MIR isn't tied to a single tool or purpose; rather it is the name for a range of solutions to a variety of digital problems.

> Music information retrieval (MIR) is a multi-disciplinary effort focused on extracting information from music and using this information to solve a wide range of problems including beat detection, automatic music transcription, artist recognition, genre classification and music recommendation. To solve these problems, researchers must develop algorithms that can extract salient musical information directly from the audio. This can be extremely difficult. (Lamere 2008: 101)

The issue is that digital technologies are creating more data and this data is only as useful as it is knowable. As archives of digital music grow, the problem of how to produce relevant knowledge about those archives also grows. So far, the answer to the problem of too much data is more data – metadata, or information about information. To do this for digital music, through MIR, means implementing automatic data-monitoring performed by virtual, algorithmic "listening."

Jean-Julian Acouturier and Elias Pampalk dated the birth of MIR research as 1996, with "the first research publication trying to access musical content in a database, based on algorithms that analyse the audio content rather than on editorial information" (2008: 87). By "editorial information," they mean human input, coding digital music files based on interpretative categories invented and analyzed by (usually expert) listeners (e.g. the Pandora internet radio system). Though searchable archives of digital music are expanding, Acouturier and Pampalk found that "[m]ost of the information is annotated manually (no automated analysis), unstructured (no taxonomy), in a collaborative, dynamical and unmoderated process (unlike a centralized library)" (2008: 87). This reserve of information is hardly worthless, however: "[t]he result is a collaborative repository of musical knowledge of a size and richness unheard of so far" (2008: 87). Surely, this repository of knowledge, much of which is created by non-experts, fans and the like, is itself valuable and a significant point of departure from editorial models relying on professional listeners. But the next step for MIR would be to use these editorial contributions to train artificially intelligent software agents to "listen" sociosentimentally, as humans do.

The new frontier of MIR research is, as Acouturier and Pampalk described, to process more than mere "surface" features, meaning the empirical aspects of the digital file itself or of the resulting audio signal. This is how applications like iPhone's Shazam, work, by using spectrographic analysis to take an "acoustic fingerprint" and match it to a library of known music. But the brass ring for MIR is "higher order" processing capable of using "contextual knowledge," namely the system of human significance, or sociosentimentality, which is nowhere evident on the surface alone. However, to program artificially intelligent listening, several problems must be solved. Efforts to classify songs automatically based on characteristics of the singer (Tsai, Rodgers, and Wang 2004) or of a melody (Clausen and Kurth 2002) run into the problem of object identification, or how to automate what comes naturally to humans: identifying which elements within a complex signal, like recorded music, are attributable to a singer's voice, as opposed to the sound of the backing music, or to one melodic instrument among many within a polyphonic arrangement. With efforts to classify digital music files through crowd-sourcing, the challenges include how to automatically analyze irony, like tagging Paris Hilton as "brutal death metal," and misclassification, like tagging The Beatles as "baroque" (Acouturier and Pampalk 2008: 90), or the interpretation of synonyms, polysemes, and misspellings (Lamere 2008). Further, all of these issues are commingled with matters of digital copy-culture, such as the potential for there to be many instantiations of a song which, as far as a human listener is concerned, are the same but, as far as a data monitor is concerned, are unique because, for example, the bit-rates are different or the file names don't match.

In all of these cases, the crux of the matter is figuring out what an artificial listener should monitor given the interests of its users and the realities of digital storage and transmission. Contending with sound and listening in a purely digital domain – where sound is data and listening is the algorithmic production of metadata – neither resolves nor precludes pre-digital problems of listening cultures, it extends them. The hope for MIR is to create an artificial intelligence to perform auditory work as well or better than human "editors." This appears to be what Michel Serres (1985) feared, namely that "true algorithms" would come to replace language (which had already replaced the sensual subject). But what's exciting, or edifying, if nothing else, is that, on the forefront of MIR, one may witness the evolution of digital protocols concerning the digital-ness of the record-object.

Extracting, measuring, ordering, and storing data leads to the problem of how to make that data meaningful. The way to do this, with MIR, is to extrapolate from human listening a model for digital music information processing. Although present visions for MIR stress automatic archive organization and digital rights management, the potential impact on music production could be enormous. Consider that success in MIR could lead to a digital mediascape where music production is guided by the effects of MIR on distribution and consumption. Music for speaker-listeners may become music for algorithmic listeners. If digital music, in order to be locatable and, therefore, heard by listeners must first be knowable to data monitors, what would matter first and foremost would be not how the music will sound to humans but how it will be recognized by algorithms modeled on listening. If acoustic fingerprinting becomes more widespread, for example, recordists might well pursue new protocols for producing data easily read by such systems.

For sensory studies, the question concerning MIR and other forms of data monitoring is not how, technically, to accomplish these goals, but how the emerging digital protocols affect the current supplementation and subversion of the human ear. Due to the social construction of historical and material circumstances in which sound may be known outside of acoustic space, listening, too, may come to reside outside of listeners. Again, if technoscientific particularity is the material basis for sociosentimental experience, then recordists should be keenly interested in ways to know the most objective, empirical details of their art, the digital code itself, as they relate to sounds with cultural resonance. That is, efforts to automate listening rest on how well a data monitoring agent can, operating solely at the level of data, produce sociosentimental knowledge, or at least relevant metadata, that, until now, could come only from human listening. Yet whereas MIR researchers envision ways to monitor digital music by making machines to listen and understand for us, there is another emerging protocol. Rather than replace listening with an automatic data monitor, listening can instead be de-centered

and diffused; the data can be monitored as speaker-music by a network of listeners.

Research on social network sites (SNS) has exploded recently; however, the emphasis has been on sites intended for the broadest user base, like Friendster, Facebook, and MySpace. Much less attention has been given to specialized social networks created exclusively for digital music production and promotion, but these are crucial to consider because they demonstrate another possibility to come from rendering sound as data, a possibility where sound and listening are returned to human observers, the listening audience, in a way that the earlier history of recording couldn't accommodate.

As far as a politics of listening is concerned, music-oriented SNS are among the most significant arenas of the digital age where one encounters active and critical listening communities connecting experts, amateurs, and fans for the purpose of evaluating digital music recordings through an organized feedback system. Demo review websites such as GarageBand, IndieCharts, or SliceThePie offer both expert opinions as well as feedback from other users in exchange for a fee or for contributing reviews. To take GarageBand as an example, the network functions as a blind peer-review system. In addition to narrative criticism, each reviewer fills out a standard form covering elements of a song's production, performance, and composition, including sociosentimental information like identifying other artists or songs reminiscent of the songs in question.

This means the kind of critical listening once left to the best guessing of music producers may be networked to address perennial issues in the encoding of music for a mass audience. For instance, if one would like to know if a "Death Metal Crunch" preset made a convincingly death metal sound, one could submit the resulting digital audio file to a network of death metal fans. While it doesn't yield a body of consistent, empirical knowledge, like visual analysis or some strains of MIR might, networked listening leads to better knowledge of a record-object simply due to its aggregation of opinions in large number, the wisdom of listening crowds. Considering the promise of SNSs to preserve and extend human listening against non-auditory threats *and* the contributions SNSs could make to sociosentimental knowledge about music, networked listening ought to receive greater attention in future scholarship. Additionally, networked listening, in light of the previous analysis of isolative monitoring protocols, should be recognized as a major techoustemological shift away from protocols treating the record-object as a thing beyond listening and towards protocols that may reintegrate it into the world of sensuous reality as an object for listening through digital networks.

While better understanding of listening SNSs requires further study, the existence of these networks highlights the fact that – thanks to the convergence of multiple media through personal computers and the user-friendliness of graphic interfaces, visual analysis, and so on – digital audio technologies de-skill listening

and, therefore, expand opportunities for people to participate in the production, distribution, and criticism of music recordings. But the trouble with increased participation is the already vast and ever expanding reserve of raw, digital information – data – that needs to be identified, organized, and tracked, which strengthens the drive to subvert listening entirely by improving automatic data monitoring like MIR.

Conclusion

Digital audio technologies enable widespread participation in record production and consumption, but they also de-skill, fragment, and displace human listening. New ways of monitoring digital music aim to produce knowledge beyond listening because listening alone cannot meet the demands of a mass-mediated, digital audio culture. Protocols of audio professionals, prior to the digital age, produced an object of listening that could defy listening itself: the record-object. The social construction of the record-object came with the possibility of recording music, which problematized the relationship between listeners and the object of listening in ways both empirical and cultural. Examining audio, visual, and data monitoring shows different protocols for knowing recorded music, demonstrating a trend toward producing more information about sound while reducing the amount of critical listening needed to produce and understand that information.

Extending the realm of audile knowledge into the visual shows that it's possible to supplement listening with visualization technologies, and even program it, to an extent, through open mixes and presets. The digital format also enables networked listening, an emerging monitoring protocol for knowing how sounds work in their intended social contexts. MIR attempts to go one step further, toward rendering audile skills as artificially intelligent "listeners" designed to process data into something more significant and productive of further knowledge appropriate and necessary for sustaining and controlling digital culture.

Acknowledging that music is made with regard to knowledge about the conditions of its circulation and reception, one must now be attentive to the conditions affecting a new techoustemology of digital music. Observing these conditions, key areas for future research include: (1) visualization and new terrains of multisensory knowledge – such as how sight and sound work together to broaden participation and expedite the production process; (2) data analysis – especially algorithmic listening to produce metadata about audio files; and (3) the material and social environments for audience reception and feedback – including trends in consumer audio equipment and the use of SNSs to organize critical listening communities.

Acknowledgments
Special thanks to Jonathan Sterne for feedback on earlier drafts of this essay. Thanks also to Adam Roth and Suellen Adams for opportunities to present this work while under development, and to the reviewers at *The Senses and Society* for valuable help bringing to fruition this final version.

Notes

1. Non-auditory technologies are *about* sound and listening but rely on them only in the final instance.
2. After W.J.T. Mitchell's argument that the declining epistemological status of speech and writing in the age of the image was a "sure sign that a pictorial turn is taking place" (1994: 13), it is now arguable that another, digital turn is taking place, signaled by the way digital audio processes having little to do with human auditors are displacing older, audile and even visual processes.
3. This term is borrowed from Lisa Gitelman's (2006) work wherein "protocol" means, basically, technique developed within a larger power structure. In sensory studies, this is like Caroline Jones' modernist "regime" of vision and the "bureaucratizing" of the senses supporting "the visibility" (2005). Monitoring protocols are part of a sensory regime supporting "the audibility" of digital music.
4. This de-skilling can be found at other pivotal moments in music cultures. Myles Jackson (2008) and Trevor Pinch and Frank Trocco (2004) identified similar moments. In each, expert listeners once required for the care and maintenance of highly technical apparatuses – orchestras and organs in Jackson, synthesizers in Pinch and Trocco – were replaced by tools, like tuning-forks and pre-wired patches, automating the knowledge of audio experts.

References

Acouturier, Jean-Julian and Pampalk, Elias. 2008. "From Genres to Tags: A Little History of Music Information Retrieval." *Journal of New Music Research* 37: 87–92.

Berger, Harris and Fales, Cornelia. 2005. "'Heaviness' in the Perception of Heavy Metal Guitar Timbres: The Match of Perceptual and Acoustic Features Over Time." In P.D. Greene and T. Porcello (eds), *Wired for Sound: Engineering and Technologies in Sonic Cultures*. Middletown, CT: Wesleyan University Press.

Clausen, Michael and Kurth, Frank. 2002. "A Unified Approach to Content-Based and Fault-Tolerant Music Recognition." *IEEE Transactions on Multimedia* 20: 100–29.

Crary, Jonathan. 1992. *Techniques of the Observer: On Vision and Modernity in the Nineteenth Century*. Cambridge, MA: MIT Press.

Gitelman, Lisa. 2006. *Always Already New: New Media, History, and the Data of Culture*. Cambridge, MA: MIT Press.

Hall, Stuart. 1999. "Encoding, Decoding." In S. During (ed.), *The Cultural Studies Reader*. New York: Routledge.

Hennion, Antoine. 1990. "The Production of Success: An Anti-Musicology of the Pop Song." In S. Frith and A. Goodwin (eds), *On Record: Rock, Pop, and the Written Word*. New York: Routledge.

Jackson, Myles. 2008. *Harmonious Triads: Physicists, Musicians, and Instrument Makers in Nineteenth-Century Germany*. Cambridge, MA: MIT Press.

Jones, Caroline. 2005. *Eyesight Alone: Clement Greenberg's Modernism and the Bureaucratization of the Senses*. Chicago: University of Chicago Press.

Jones, Steve. 1992. *Rock Formation: Music, Technology, and Mass Communication*. London: Sage.

Karalian, Alan. 2009. Letter to the Editor. *Tape Op: The Creative Music Recording Magazine* 73: 14.

Lamere, Paul. 2008. "Social Tagging and Music Information Retrieval." *Journal of New Music Research* 37: 101–14.

Lehrman, Paul. 2005. "Do You Hear What I Hear? Learning to Listen in a Mediated World." *Mix*; available online: http://mixonline.com/mag/audio_hear_hear/index.html.

Manning, Peter. 1993. *Electronic and Computer Music.* New York: Oxford University Press.

Mitchell, W.J.T. 1994. *Picture Theory*. Chicago: University of Chicago Press.

Owsinski, Bobby. 2006. *The Mixing Engineer's Handbook*. Vallejo, CA: Mix.

Pinch, Trevor and Trocco, Frank. 2004. *Analog Days: The Invention and Impact of the Moog Synthesizer*. Cambridge, MA: Harvard University Press.

Porcello, Thomas. 2005. "Afterword." In P.D. Greene and T. Porcello (eds), *Wired for Sound: Engineering and Technologies in Sonic Cultures*. Middletown, CT: Wesleyan University Press.

Serres, Michel. 1985. *The Five Senses: A Philosophy of Mingled Bodies*. Translated by M. Sankey and P. Cowley. New York: Continuum.

Sterne, Jonathan. 2006. "The Life and Death of Digital Audio." *Interdisciplinary Science Reviews* 31(4): 338–48.

Tsai, Wei-Ho, Rodgers, Dwight and Wang, Hsin-Min. 2004. "Blind Clustering of Popular Music Recordings Based on Singer Voice Characteristics." *Computer Music Journal* 28: 68–78.

Zak, Albin. 2001. *The Poetics of Rock*. Berkeley, CA: University of California Press.