Louisiana Tech University

From the SelectedWorks of Michael Austin

2016

“Safe and Sound: Using Audio to Communicate Comfort, Safety, and Familiarity in Digital Media"

Michael Austin, Ph.D., *Louisiana Tech University*

Available at: https://works.bepress.com/michael-austin/1/

CHAPTER 2

Safe and Sound: Using Audio to Communicate Comfort, Safety, and Familiarity in Digital Media

Michael L. Austin
Howard University, Washington, DC, USA

As I constantly searched for a quiet place in which I could sit to write this essay, one fact became increasingly clear: sounds are *everywhere*. And understandably so, especially since sounds perform an important function in how we receive information from our environment—even expressive, emotional information. Sound has been used for centuries as both a way to express emotions and as a method to elicit emotional responses from listeners; from audible speech and visceral utterances, to expressive music in concerts, film, and theater, sound has always played a major role in affective interactions. Many everyday objects are constructed to make sounds with affective properties and are increasingly designed and utilized to communicate or elicit positive emotional responses in consumers of digital media. This is particularly true for objects that provide a sense of comfort or safety, as they signal successful user experience (UX) design.

For example, imagine stepping out of your car or a taxi and closing the door behind you as you hurry off to a meeting. Your visual imagination can probably conjure up many details regarding the high level of design that went into the manufacturing of your car, especially if it is a luxury car, such as BMW: sleek, aerodynamic contours, an idiosyncratic shade of black with metal flecks in the paint, fashionable leather seats, and a specially designed control interface with icons and fonts exclusive to BMWs. Just as design features convey a sense of style and luxury, the snug feeling of the seatbelts, the quick response of the breaks, and the digital screen that allows the driver to see behind the car are all designed to express a sense of safety. Within your aural imagination, can you hear the sound of the car door slamming shut? If this car is, in fact, a BMW, there is a good chance that the sound the door

Emotions, Technology, and Design
http://dx.doi.org/10.1016/B978-0-12-801872-9.00002-8
© 2016 Elsevier Inc.
All rights reserved.

makes as it closes was designed by Emar Vegt, an "aural designer" who works at the company's head office in Munich. Describing this design element in an interview with David Baker for *Wired Magazine*, Vegt says, "The sound of the door closing is a remarkable aspect of the buying decision … It gives people reassurance if the door feels solid and safe" (Baker, 2013). Likewise, other sounds in BMWs are designed to be a little discomfiting as a reminder to think about safety or to account for the safety of others. Baker goes on to write:

> *Inside the car are other considerations. "Warning sounds need a particular aesthetic," he says [citing Vegt]. The noise that tells the driver to put on their seat belt can't be too pleasant as "people will listen to it like a symphony." But neither can they be too annoying—people find ways of shutting them off. And electric cars are a challenge. "Sounds tells [sic] people that a car is there, which is really important for blind people."[1]*
>
> ***(Baker, 2013)***

In order to assure their car sounds the way it should, "every sound made by a BMW is analyzed by a team of over 200 acoustic engineers to ensure they are both mechanically and acoustically correct" (Jackson, 2003, p. 106).

Just as the closing of a car door is carefully designed to give the driver the feeling it has been safely shut and adds to a sense of confidence in the entire car, designers of audio for digital media often want to educe feelings of security and familiarity as this signals successful UX design. In this essay, I explore significant uses of sound and audio feedback to communicate feelings of comfort, security, and intimacy in digital media. Designers—sound designers, UX designers, human-computer interaction (HCI) designers, interface designers, and designers and engineers from a host of disciplines—often seek to create positive, engaging experiences with technology, and do so through the utilization of sonic material to ensure that a user's encounter with technology is a pleasant one. Using case studies involving interfaces designed to communicate safety and comfort, I frame listening, particularly semantic modes of listening, as the primary way in which the emotional information conveyed through digital media is understood,

[1] Here, Vegt is making reference to the fact that electric cars were found to be more dangerous to pedestrians than gas-powered cars because they were much quieter, almost silent, and the visually impaired were sometimes unaware of their presence. To address this problem, the European Parliament ruled that electric and hybrid cars must add artificial noise to their engines. See Walker, A. "Silent But Deadly: The EU Wants Electric Cars to Add Sounds for Safety." Gizmodo, April 7, 2014. http://gizmodo.com/silent-but-deadly-the-eu-wants-electric-cars-to-add-so-1560215281.

and I present rationale for considering listening, hearing, and the sounds themselves as equally necessary parts for an emotional understanding of digital media.

Before digging deeper into more examples of the possible ways in which this can be accomplished in digital media, primarily through auditory display, I first want to examine auditory awareness (i.e., the ways in which we are attuned to our sonic environments) in order to help frame an understanding of the ways in which one might listen to sounds from digital media within our environment, and how one might derive emotional meaning from them.

HEARING, LISTENING, FEELING

Aural information gathering, especially emotional information, does not simply depend upon a listener's ability to perceive sound as it propagates throughout an environment and process it physiologically. Rather, the ability to hear a sound physically, process it semantically, and understand it semiotically necessitates a broader understanding of the interconnected relationships of sounds, the way(s) we listen to them, and the environment from which they emanate. In *Background Noise: Perspectives on Sound Art*, Brandon LaBelle writes:

> *Sound is intrinsically and unignorably relational: it emanates, propagates, communicates, vibrates, and agitates; it leaves a body and enters others; it binds and unhinges, harmonizes and traumatizes; it sends the body moving the mind dreaming, the air oscillating. It seemingly eludes definition, while having a profound effect.*
>
> ***(LaBelle, 2008, p. ix)***

We use sound to connect with others emotionally. Even before we are born, we are able to hear our mother's voice, distinguish it from other voices, and react to it (Kisilevsky et al., 2003). Lullabies are sung to infants and young children to help them feel connected to the singer and to give them a sense of safety and comfort. Studies indicate that music can help children relax, diminish their pain, and reduce their anxiety (Longhi, Pickett, & Hargreaves, 2015).

Clearly, sound has an impact on how we feel, but how are vibrations of air molecules able to bring us to tears, strike terror in our hearts, or make us feel safe? Throughout our lives, sound facilitates the expression of emotional communication with others; our ability to perceive any sonic information relies on our ability to process the auditory information we collect from our environment. Researchers in human hearing, psychoacoustics, and cognition have identified several major skills that comprise healthy auditory processing. According to the American Speech-Language-Hearing Association, these skills are:

- Sound localization and lateralization (identifying the *place* of sound).
- Auditory discrimination (identifying different sounds and the differences among them).
- Auditory pattern recognition (identifying patterns in iterations of sounds, including those identified within Gestalt theory, such as grouping, figure/ground, good continuation, expectancy, etc.).
- Temporal aspects of audition, including:
  - temporal resolution (detecting a rapid succession of consecutive sounds as separate, rather than a single sonic event),
  - temporal masking (a process wherein sudden changes in the volume of a sound can "mask," or hide, the sonic event preceding or following it from our hearing),
  - temporal integration (combining patterns of sounds, or recognizing the contours that comprise a sound's envelope—attack, decay, sustain, or release—and translating that into useful information), and
  - temporal ordering (distinguishing the order in which successive sounds are heard).
- Auditory performance in competing acoustic signals (including dichotic listening; the ability to focus attention on important sounds and ignore background noise)[2] and
- Auditory performance with degraded acoustic signals.

(American Speech-Language Hearing Association, 1996, *parenthetical definitions mine*).

While auditory processing is a multimodal set of purely physiological activities, understanding the emotional content in sonic information, particularly in speech, relies heavily on these processes, especially since many of the ways in which emotions are communicated is through sound, such as tone of voice, sighs, tempo and volume of speech, vocal cadence, etc. Further, not only does central auditory processing disorder cause those with it to struggle with one or several of the skills listed above, resulting in difficulty understanding spoken communication, it can also possibly lead to a misunderstanding or misreading of the emotional cues embedded within speech.

[2] Also known as the "cocktail party effect."

The large and growing body of literature in psychological research on the affective/emotional qualities of music could also have implications for this examination of the emotional impact of sounds (both musical and otherwise) in digital media. Research in this area has been conducted for well over a century; recently, the approaches taken have either been to investigate possible causes of emotional arousal through music, or to investigate ways in which music might mediate the experience of emotions. Among causal factors, scholars list the listener's age, sex, music education, the physical environment wherein music is experienced, and whether or not the listener was alone or a member of an audience (Abeles & Chung, 1996; Gabrielsson, 2001), and some even question if there is something intrinsic within the music itself that elicits emotions (Sloboda, 1991). Others describe the underlying mechanisms that induce emotions, such as cognitive appraisal (Scherer, 1999), musical expectancy—(i.e., whether or not the music confirmed or defied listener's expectations (Meyer, 1956)), mental images that music could possibly evoke (Lyman & Waters, 1989; Osborne, 1980; Plutchik, 1984), and even brain stem reflexes and episodic memory (Juslin & Västfjäll, 2008). Although these and other approaches have not yet resulted in a single satisfactory conclusion regarding the source of music's emotional power, perhaps it accounts for some of the ways in which the same musical work can affect a wide array of seemingly contradictory emotions. As with auditory processing, these psychological approaches relate to the ways in which our ears perceive physical sound (i.e., how we actually *hear*), and possible ways in which our brains can perceive aural information and interpret it in an emotional manner.

Listening can, of course, be approached from more semiotic, philosophical, political, and theoretical perspectives as well, especially when considering our mutable relationship with sound. Because sound plays an important role in the way we experience the world, it also plays an equally vital role in the way we approach and begin to understand it. The most obvious connection between the philosophical and the ecological is made in the work of ethnomusicologist Steven Feld; he coined the term *acoustemology* (or acoustic epistemology), defined as "local conditions of acoustic sensation, knowledge, and imagination embodied in the culturally particular sense of place" (Feld, 1996, p. 91). In other words, one can understand the surrounding world, epistemologically, through its sounds. Pointing toward similar conclusions, Jacques Attali began his now-famous treatise, *Noise: The Political Economy of Music*, writing:

> *For twenty-five centuries, Western knowledge has tried to look upon the world. It has failed to understand that the world is not for the beholding. It is for the hearing. It is not legible, but audible. Our science has always desired to monitor, measure, abstract, and castrate meaning, forgetting that life is full of noise and death alone is silent … Now we must learn to judge a society more by its sounds …*
>
> **(Attali, 1985, p. 3)**

We can evaluate a culture and its values by examining its acoustic culture. In examining the sounds produced by our society, we begin to assign value to what we hear, most notably in the way we categorize the aural: silence versus sound versus noise, good versus bad, music versus Muzak, and eventually more emotional dichotomies such as pleasure versus annoyance, or even happy versus sad. Describing the ways in which categorizing the aural is a facet of social life, Ian Biddle writes:

> *As a system by which the conceptual territories noise/music/silence are mapped and managed, the political ontology of sound is also a political theory of relationships: there is no quiet without less quiet, no noisy without less noisy, no music without its forbidden others. Class, ideology, race and gender are all visitors to this process of naming, of holding apart, and holding in mutually exclusive relation the three territories. They all make their way, like a little tiny parasitic relation of their own, into the mechanisms by which noise-obsessed neighbors, anxious public license granters, social theorists and policy makers seek to discipline and silence the social.*
>
> **(Biddle, par 2)**

These attributes of acoustic culture correspond to the causal factors in previously mentioned research in music psychology, such as the way in which a listener's environment can possibly affect the mood or emotions that are elicited by music experienced in a particular time and place. Not only are these location-based causal factors used to contextualize musical and nonmusical sound, but they also present an ecological frame of reference for the emotional arousal triggered by them.

Thus far, I have described just a few of the myriad ways in which scholars, theorists, and researchers have attempted to label specific ways in which we attend to sound and account for emotional responses that result from this hearing. The sum of these parts that constitute the multimodality of hearing is listening, that is, *hearing to understand*. And just as there are many modes of hearing, listening, too, comes in many varieties. Pierre Schaeffer, a composer and one of the pioneers of the *musique concrète*[3] tradition, developed four

[3] Concrete music; this is a compositional practice wherein composers took both electronically-produced and found sounds, recorded them, and then manipulated them in an attempt to make them *acousmatic* (simply, sound that is heard but the listener cannot immediately identify its source).

modes of listening, or *Quatre Écoutes*, to illustrate the levels at which hearing, listening, and understanding are interrelated (Schaeffer, 1977). These levels are:

- *Ouïr:* This mode simply describes passive hearing. Sound waves strike our ears and we hear them, yet sometimes we do not seek to comprehend them. This applies to background noise and other environmental sounds that do not demand attention.
- *Écouter:* This is the most basic form of listening wherein we hear a sound and pay attention to it. Here, we consider sound in more semiological terms (What is the source of this sound? What does it indicate?).
- *Entendre:* This mode implies listening with intent; it is a combination of *ouïr* (an objective, physical hearing) and écouter (an objective, semiotic hearing); herein, a listener subjectively chooses the sound to which he or she pays attention.
- *Comprendre:* In this subjective mode, the listener prioritizes sound, decides which sounds are significant or irrelevant, and assigns meaning to sound. Here the sound object's essence is rendered irrelevant (i.e., What is special about this particular sound?), and that which is represented by the sound comes to the forefront (i.e., What does the sound *mean*? What does it *represent*?).

These four modes of listening are famously summarized by Schaeffer in one sentence: "I heard (ouïr) you despite myself, although I did not listen (écouter) at the door, but I didn't understand (comprendre) what I heard (entendre)" (Paraphrase of Chion, 1983). While most listening involves all four modes at once, it is at the *comprendre* level that the emotional qualities are assigned to sounds, derived from sounds, or represented by sounds. Michel Chion combines these listening modes into three:

- Causal Listening: "consists of listening to a sound in order to gather information about its cause (or source)"(Chion, 1990, p. 25).
- Semantic Listening: this involves listening to the "codes" within a sound and finding the meaning therein.
- Reduced Listening: "takes the sound-verbal, played on an instrument, noise or whatever—as itself the object to be observed instead of as a vehicle for something else" (Chion 1990, p. 29). This is a phenomenological approach to the experience of sound that reduces a sound to its essence and investigates its qualities (timbre and harmonic content, form, volume, etc.).

Chion's "semantic listening" relates most to Schaeffer's "comprendre," and it is this listening mode with which we decipher the emotional meaning encoded within a sound. Expanding on David Huron's six-component theory of auditory-evoked emotion, Kai Tuuri, Manne-Sakari Mustonen, and Antti Pirhonen devised a system to understand the combination of the psychological and the philosophical modes of listening:

Preconscious modes:

- Reflexive: reflexive responses triggered by sound.
- Connotative: freely formed associations immediately evoked in listening.

Source-orientated modes:

- Causal: listening for the cause of a sound.
- Empathetic: listening for emotion or state of mind expressed through sound.
- Critical: critiquing a sounds suitability for a particular situation.

Quality-oriented mode:

- Reduced: objectively describing the properties of a sound.

In order for sound in digital media to make an emotional impact, an awareness of both the psychological and physiological immediacy of auditory awareness, and the philosophical, semiotic modes of listening and understanding the possible meanings conveyed with sounds, is required. From a design standpoint, engineers, programmers, and other creators of digital media keep both viewpoints in mind when creating sounds with emotional functions and their corresponding visual and haptic counterparts. Designing effective auditory feedback requires an understanding of auditory processing and the perceptual limitations of sound perception for users; for example, a designer might ask: "Is the sound loud enough for the user to hear? Is the pitch too low or too high to easily hear? Are important feedback sounds somehow hidden by louder, more trivial mechanical sounds produced by the device in question? Are sounds emitted too close to one another, causing them to be temporally masked?" Likewise, designers must also ask deeper philosophical questions: "What do the sounds of these instruments represent? Do they represent the same thing in every culture? At what point does a sound become noise, and when does a sound pass the point of being an alert to being a nuisance? Can a user easily derive the meaning from this sonic feedback alone, or are other sensory cues (visual, haptic, etc.) required?"

In his book, *Emotional Design: Why We Love (or Hate) Everyday Things*, Donald A. Norman discusses his research on the emotional impact of esthetics and design, concluding that emotions and cognition are intertwined. According to Norman, when a user encounters a designed product, we process our experience through three levels of perception: *visceral, behavioral, and reflective.* A user first experiences design on a visceral level, which includes affective reactions and emotions, and this is the most primal and reactionary form of perception wherein we decide if the design is good or bad, pleasing or disgusting, safe and comforting, or dangerous, or alienating, etc. This level of experience could correspond to Schaeffer's "*oüir*" and/or "*écouter*" modes of listening, to Chion's "causal listening," and to Tuuri, Mustonen, and Pirhonen's preconscious modes of listening. Experience of design on the behavioral level is triggered by our reactions on the visceral level, and like the visceral level, is mostly unconscious; it describes how users act and feel in using the design, and whether or not using the design creates a meaningful experience, corresponding to Schaeffer's *entendre* and Tuuri, Mustonen, and Pirhonen's "causal" mode of listening. Last, reflective processing refers to the cognitive processing of a design and the rationalization of choices made as a result of the design. In this level of processing, users seek to understand a product, assess its value or ascribe value to it, even to the point of integrating the design into the expression of self-image, to have pride in the ownership of the product, and to attribute some cultural value to the design. This reflective level parallels Schaeffer's "*comprendre*," Chion's "semantic listening," and Tuuri, Mustonen, and Pirhonen's "empathetic" and "critical" modes of listening.

Our experience with a designed product begins with visceral and affective processing. Our first experience with design is marked by our emotional reaction to it, and our continued interaction depends upon whether or not the design affected an appropriate, desirable emotional response. Sound can reinforce or negate the emotional content designed into the visual elements of digital media, affecting our behavior. For example, a love song that plays in the background as star-crossed lovers meet for the first time makes a movie scene all the more saccharine, resulting in more emotional engagement, tears, etc.; likewise, horror movies often employ anempathetic sounds or music that incongruously signals stasis, calmness, or even happiness while a brutal murder is occurring within the visual *mis-en-scène* and is exploited to produce an additional feeling of uneasiness. Similarly, smartphone games with repetitive music or applications that produce sonic alerts that are too loud or those that occur too often are frequently countered with the phone's mute button. Upon hearing a particular ringtone, listeners can exercise their reflective perception, making value judgments about the brand of phone that plays such a ringtone, about the person who would own that brand of phone, and even compare this evaluation to past experiences with the same ringtone—it can remind a listener of happier times because this particular ringtone plays each time a loved one calls, or it can be an annoyance to an innocent bystander who is distracted from his or her work a little more each time the ringtone resounds loudly across the office.

(AUDITORY) DISPLAYS OF EMOTION

Rather than simply relying on one sense at a time, we interact with the world using many senses simultaneously. The combination of visual and aural feedback in digital media gives us a great deal of information, especially at the site of the human-computer interface. In an essay on nonspeech auditory output, Stephen Brewster lists several reasons why sound is beneficial in HCI (Brewster, 2007, p. 249). Vision and hearing are interdependent, and our ears signal to our eyes that there is something that demands visual attention; the temporal resolution of our auditory system is superior to that of our visual perception. Also, sound reduces the overload from large displays, so rather than bombarding users with tons of visual information that can easily be overlooked, some of that information is presented as sound instead. Sound also reduces the amount of information needed on screen and reduces the demands on a user's visual attention. While attending to a task that requires visual attention, one can rely on aural feedback to monitor the progress of other; for example, downloads and other more time-intensive processes are often assigned a sound that users hear whenever the process is complete. Meanwhile the user can attend to other business and wait for the signal, rather than constantly checking to visually confirm whether or not the download is complete. Similarly, when buying groceries at the supermarket, the cashier does not need to continually check the screen on the cash register to make sure an item was properly scanned; rather, the "beep" emitted from the machine provides aural confirmation that the item is accounted for, and the cashier can quickly continue to sell groceries. Brewster also notes that the auditory sense is underutilized, that sound is attention grabbing, that some objects or actions within an interface may have a more natural representation in sound, and that making computers more usable by the visually impaired is among the benefits of using sound in HCI design contexts[4] (Brewster, 2007, p. 249). While Brewster does not mention the emotional connections and connotations associated with sound and music, these are certainly assets that sound brings to HCI and digital media.

[4] Quoting G. Kramer, Brewster also lists some problems with using sound in HCI: the relatively low resolution of sound compared to high resolution visuals, presenting absolute data with sound is difficult, there is a lack of orthogonality and changing one attribute of one sound could affect the others, sound is temporal and information represented with sound is transient, and finally, some people easily find sonic feedback annoying (Brewster, 2007, p. 249).

Auditory display is the general term used to describe the ways in which sonic feedback from an interface communicates information to the user in HCI. The process of representing or perceptualizing data or other information using sound is called *sonification*. Reading printed information orally is an easily recognizable example of sonification—the information on the page is transduced from print (perceived visually) into intelligible speech (perceived aurally). Auditory displays are used within an interface to perform a number of functions, all of which communicate information to the user. According to Bruce Walker and Michael Nees (and many others including Buxton, 1989; Edworthy, 1998; Kramer, 1994), auditory displays perform alerting functions (esp. notifications and warnings), status and progress indication functions, data exploration functions (which is most applicable in the sonification of scientific data), and in art and entertainment applications (for the creation of computer music, digital sound art, immersive exhibitions, games, etc.) (Walker & Nees, 2012, p. 4).

Auditory displays employ several types of auditory feedback that are effective intermediaries of emotional information. *Auditory icons*, an idea developed by Gaver (1989, 1997), are the aural equivalent of visual icons and rely on their identifiablity from the analog world to transfer meaning to the digital domain. Visual icons are part of the graphical user interface that helps users to understand and interact with an electronic device; for example, users recognize the image of a trashcan on their computer desktop and understand that dragging a file to this icon presumably results in "trashing" the file, essentially deleting it, and freeing space on the computer's hard drive. The sound of crumpling paper the user hears as he or she drags a file into the trashcan icon is an auditory icon—it represents the act of discarding a file similar to the visual depiction of the trashcan.[5] Any emotional connection to a real-world sound can presumably transfer into HCI contexts if the sound is used as an auditory icon. Some auditory icons, such as the "clicks" that suggest a user is typing a text message on a smartphone, or the shutter sound that is used to indicate a picture is taken with a smartphone camera, are skeuomorphic[6] and are not necessary, per se (as any other sound could

5 On computers that run Windows operating systems, the trashcan is exchanged for a more ecologically friendly "recycling bin." When it is utilized, users can hear the sound of a crushed aluminum can being thrown into an empty bin.

6 A skeuomorph is a design element, usually ornamental, that mimics design qualities of older versions; digital skeuomorphs mimic design features of physical/analog objects. The tiny ring on the neck of some syrup bottles, the metal rivets on jeans, and the turning pages of e-books are all skeuomorphs.

replace them, or they could still function just as well with no sound at all), but they give users aural feedback that the interface is responding to his or her input, resulting in ease of use, positive action reinforcement, and ultimately results in positive emotional interactions with the interface for the user. In fact, skeuomorphic auditory icons are often used in digital media with the hope to elicit positive feelings of nostalgia in users. The camera shutter click on their camera phone reminds the user of the good times he or she had with their analog camera, the sound of a record ending that is artificially added to the end of some CDs hearkens back to the "good old days" of long play records (LPs), and users can choose ringtones that sound like an old-fashioned telephone for old times' sake. Apple iPhones and other smartphones also employ auditory icons that signify safety and security; when the phones are locked, users are reassured by the sound of a clicking lock that their data is secure. Digital security systems also reassure users that the system is armed and that their house is satisfactorily protected through voiced announcements or simple beeps.

"*Earcons*" are similar to auditory icons, but rather than being concrete, real-world sounds transposed into a digital interface, earcons are abstract, sometimes musical in nature, and are manufactured to present or represent information aurally to the user. Sonic branding and sonic logos (sometimes called "sogos") are perhaps the most famous types of earcons; in the USA, the three-tone NBC sonic logo, Intel's 5-tone sonic logo, the chord an Apple computer plays as it starts up, all instantly represent their corresponding brand to the listener. Unlike auditory icons, however, earcons are not everyday sounds that are mapped onto functions of the interface, so rather than being intuitively understood, the meaning of earcons must be learned with the system. Emotional responses to earcons are forged by emotional connections to brands, and hearing these earcons can remind listeners of this connection.

Earcons can easily be found on personal computers; each brand of computer has its own idiosyncratic (and trademarked) set of earcons that are mapped onto a wide array of tasks. The Windows operating system has a number of earcons specific only to Windows; some are replaced with each new version of Windows, and others remain for several software generations. Each version has an identifiable, musical start-up and shutdown earcon, reassuring users that the computer has been turned on and off safely.

Windows also uses a "tada" fanfare that signals completed tasks, a sound similar to a bubble popping that is played to indicate the "popping up" of a notice of various sorts, and "dings" and error alerts (an interval of a perfect 5th played in the upper registers of a piano) notify the user something is wrong or demands their attention. These sounds help to trace out acceptable parameters for computer usage, and they afford users a sense of familiarity and offer encouraging reinforcement when heard.

Simple *speech* is also used as auditory feedback in some HCI. America online (AOL) users were greeted with "You've got mail!" whenever they received an e-mail. Many phone systems and other voice user interfaces (VUIs) afford various levels of interaction with users in the form of spoken, aural information or a series of prompts, and he or she responds vocally or with the telephone keypad. While some companies rely on recordings of real-life speakers, others rely on synthetic samples of speech. In both cases, characteristically pleasant and calming voices are chosen for their esthetic and affective qualities. Similarly, "*spearcons*," a *portmanteau* of "speech" and "icon," were created by Bruce Walker, Amanda Nance, and Jeffrey Lindsay by speeding up a spoken phrase until it is not recognizable as speech; in testing their usefulness, the creators claim that "spearcons and speech-only both led to faster and more accurate menu navigation than auditory icons and hierarchical earcons … These results suggest that spearcons are more effective than previous auditory cues in menu-based interfaces, and may lead to better performance and accuracy, as well as more flexible menu structures" (Walker, Nance, & Lindsay, 2006, p. 63). Of course, speech can be used to convey emotion, and engineers are perpetually working to refine synthetic voices into more responsive, "emotive" utterances.

Furthermore, ambient sound and environmental noise can lead to immersive and affective experiences in digital media, particularly in video games, virtual reality, training simulations, film, and various forms of entertainment media. In digital media, the affective almost always trumps the objective. In film, nothing is "real," but rather, "hyper-real"—even time and space. Scriptwriter and film producer Jon Boorstin writes:

> *Philosophers may theorize about subjectivity, but working filmmakers try to learn exactly what it means to say that time is flexible, a function of our inner clock. They study how long a second really is, and how short, and what makes it feel one way or the other. They know that what we see isn't really what's out there because they've learned how spatial perception varies with angle and focal length and lighting, how "true" colors are a figment of lighting, and context, and even the glass of a lens. They know there is no "real" sound but only a better or worse approximation of what our ear expects.*
>
> **(Boorstin, 1990, p. 198)**

Because our experience of this digital media is subjective, the onus lies on the director to affect each subjective experience to elicit the desired emotional response. For example, consider the sound design in film, especially the documentary genre. Since a documentary film is not a piece of journalism, there is no ethical mandate to report, record, or reproduce the sounds that were actually heard at the time the footage was shot. Filmmakers add extra sound effects and ambience, take out sounds, attenuates noises, fills silences with meaningful sounds, and insert nondiagetic mood music at critical points in their story, all in the service of creating a "hyper-real" soundscape that moves the audience emotionally. In many video games, immersion is critical, and players must feel as if they are part of the game in order to succeed. Ambient sound is added to represent the virtual world that is navigated by the game player's avatar, and the sound from the avatar's footsteps, gunfire, sword strikes, and other actions—even the character's breathing—are processed in such a way that they are made to be perceived as if those actions are taking place within that specific environment. For instance, if the avatar is in a cave, almost all sound effects are processed with copious amounts of reverb; if the avatar is outdoors on a battlefield, no reverb is added, accurately matching the acoustic conditions of real life and producing an aurally immersive experience with the hope that players will also become emotionally immersed within the game. It is important to remember, too, that the creation of these types of hyper-reality requires a high degree of fluency with other forms of digital media technology and software that are likewise pervaded with aural and visual feedback designed activate emotional responses (especially amusement/engagement and trust) in the user.

Much like the way in which electric cars were required to add sound because they were too quiet (see footnote 1), extra sound is sometimes added to an interface to give users the feeling that the system is still responsive and functioning properly. For instance, Skype calls, cellular telephones, and digital radio stations use "comfort noise," that is, extra static or other synthetic background noise to fill silences in the transmission; without this, silences sometimes cause users to believe the call is dropped and they will hang up prematurely thinking the transmission has failed and that they have been somehow disconnected. "Music on hold" is another example of this supplementary sound. Callers who are placed on hold will frequently hear "Muzak," "elevator music," recordings of classical music or jazz, or sometimes even a live feed from a terrestrial radio broadcast as they wait "on hold"; this music is supposed to give the listener a sense of assurance and comfort that he or she has not been disconnected.

"Extra" sound is also sometimes used to make technology more comfortable. "Siri" and other "intelligent personal assistants," or the vocal feedback provided by global positioning systems (GPS) navigation interfaces are all designed to be comforting and familiar voices with which users can interact. Advanced HCIs, namely artificial intelligence and advanced robotics, are sometimes eerie and disconcerting to humans, especially when the interface is humanoid. Computer voices have long been considered disturbing; HAL 9000 from Stanley Kubrick's film *2001: A Space Odyssey* (1968) spoke with an obviously unnatural, monotone, dispassionate voice—totally computer yet somehow still human. In the 1970s, Masahiro Mori developed the concept of the "Uncanny Valley," an esthetic theory that explains the human tendency to react with revulsion toward robots that were similar, but not quite identical, to humans (Mori, 1970/2012). As technology advances and robots continue to look, move, and act more like humans, this particular issue becomes ever more poignant, especially when these robots are being used for medical purposes and other critical social situations. To address this issue, sound libraries are being developed and implemented to help robots express intention and emotion in social interactions, thereby making them more familiar and less unnerving to humans. Others are conducting research to discover how far voice anthropomorphism can be stretched without causing unwanted emotional reactions in humans, and to discern what types of voices are best suited for emotional robot-human interactions (Cowan, 2014; Niculescu, Dijk van, Nijholt, & See, 2011; Riek, Rabinowitch, Chakrabarti, & Robinson, 2009). Sound is an important part of affective computing and has the potential to play a large part in bridging the uncanny valley.

CONCLUSION: SOUNDING SAFETY, SECURITY, STASIS, AND STATUS

Consumers tend only to purchase and continue to use products that cultivate positive feelings. Companies spend countless hours and untold amounts of money on focus groups to ensure potential customers feel happy using their product. Users rely on both positive and negative feedback from digital media to confirm whether or not they are using the interface properly. Sound plays a critical role in assuring users that they are utilizing digital media effectively and correctly, comforting them in the knowledge that they are safely using a product. Sound also signifies status for users of digital media and helps to formulate to various degrees an emotional sense of self-worth; whether a user is proudly staking claim over his or her environment by playing music loudly from a smartphone, bragging about a recent purchase of the newest phone by conspicuously leaving the ringtone on for all to hear, or by sharing nostalgic feelings and tastes in music through playing a music video game, the sonic elements of digital media make an emotional impact on the user. As technology marches toward more ubiquitous computing and the "Internet of Things," wherein users might simply use his or her voice to control household appliances, a car, and almost all other electronic devices, and as anticipatory computing develops, allowing your computer to fetch information you need before you even request it, it is imperative that the emotional impact of the sounds associated with this technology, and the ways in which emotion is communicated through these sounds, is better understood.

REFERENCES

Abeles, H. F., & Chung, J. W. (1996). Responses to music. In D. A. Hodges (Ed.), *Handbook of music psychology* (2nd ed., pp. 285–342). San Antonio, TX: IMR Press.

American Speech-Language Hearing Association. (1996). Central auditory processing: Current status of research and implications for clinical practice. *American Journal of Audiology, 5*(2), 41–54.

Attali, J. (1985). *Noise: The political economy of music.* Minneapolis, MN: University of Minnesota Press.

Baker, D. (2013). Did you know BMW's door click had a composer? It's Emar Vegt, an aural designer. *Wired Magazine*, Retrieved from http://www.wired.co.uk/magazine/archive/2013/04/start/music-to-drive-to.

Boorstin, J. (1990). *Making movies work: Thinking like a filmmaker.* Los Angeles: Salman-James Press.

Brewster, S. (2007). Nonspeech auditory output. In A. Sears & J. Jacko (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications.* (2nd ed.). Boca Raton, FL: CRC Press.

Buxton, W. (1989). Introduction to this special issue on nonspeech audio. *Human-Computer Interaction, 4*, 1–9.

Chion, M. (1983). *Guide des object sonores, Pierre Schaeffere et la researche musicale* (J. Dack and C. North, Trans., 1995). Paris: Ina-GRM/Buchet-Chastel.

Chion, M. (1990). *L'audio-vision. Paris: Nathan [Audio-vision. Sound on Screen].* New York: Columbia University Press. C. Gorbman, trans.

Cowan, B. R. (2014). Understanding speech and language interactions in HCI: The importance of theory-based human-human dialog research. In *Designing speech and language interactions workshop, ACM conference on human factors in computing systems, CHI 2014.* Available online, http://www.cs.bham.ac.uk/~cowanbr/understanding%20speech%20and%20language%20interactions%20in%20HCI.pdf.

Edworthy, J. (1998). Does sound help us to work better with machines? A commentary on Rautenberg's paper 'about the importance of auditory alarms during the operation of a plant simulator. *Interacting with Computers, 10*, 401–409.

Feld, S. (1996). Waterfalls of song: An acoustemology of place resounding in Bosavi, Papua New Guinea. In Steven Feld & Keith Basso (Eds.), *Sense of place.* Santa Fe, NM: School of American Research Press.

Gabrielsson, A. (2001). Emotions in strong experiences with music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 431–449). Oxford: Oxford University Press.

Gaver, W. (1989). The sonicfinder: An interface that uses auditory icons. *Human Computer Interaction, 4*(1), 67–94.
Gaver, W. (1997). Auditory interfaces. In M. Helander, T. Landauer, & P. Prabhu (Eds.), *Handbook of human-computer interaction* (2nd ed., pp. 1003–1042). Elsevier: Amsterdam.
Jackson, D. M. (2003). *Sonic branding: An introduction*. New York: Palgrave Macmillan.
Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences, 31*(5), 559–575.
Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H.-H., et al. (2003). Effects of experiences on fetal voice recognition. *Psychological Science, 14*(3), 220–224.
Kramer, G. (1994). An introduction to auditory display. In G. Kramer (Ed.), *Auditory display: Sonification, audification, and auditory interfaces* (pp. 1–78). Reading, MA: Addison Wesley.
LaBelle, B. (2008). *Background noise: Perspectives on sound art*. New York: Continuum.
Longhi, E., Pickett, N., & Hargreaves, D. J. (2015). Wellbeing and hospitalized children: Can music help? *Psychology of music, 43*, 188–196.
Lyman, B., & Waters, J. C. (1989). Patterns of imagery in various emotions. *Journal of Mental Imagery, 13*, 63–74.
Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago, IL: University of Chicago Press.
Mori, M. (1970/2012). The uncanny valley (K.F. MacDorman and N. Kaggeki, Trans.). *IEEE Robotics and Automation Magazine, 19*(2), 98–100.
Niculescu, A., Dijk van, B., Nijholt, A., & See, S. L. (2011). The influence of voice pitch on the evaluation of a social robot receptionist. In *Proceedings of the second IEEE international conference on user science and engineering, I-USEr2011, 29 November–2 December 2011, Kuala Lumpur, Malaysia* (pp. 18–23).
Osborne, J. W. (1980). The mapping of thoughts, emotions, sensations, and images as responses to music. *Journal of Mental Imagery, 5*, 133–136.
Plutchik, R. (1984). Emotions and imagery. *Journal of Mental Imagery, 8*, 105–111.
Riek, L. D., Rabinowitch, T., Chakrabarti, B., & Robinson, P. (2009). Empathizing with robots: Fellow feeling along the anthropomorphic spectrum. In *Proceedings of the IEEE international conference on affective computing and intelligent interaction (ACII '09), Amsterdam, Netherlands, September 10–12, 2009.*
Schaeffer, P. (1977). *Traité des objets musicaux*. Paris: Le Seuil.
Scherer, K. R. (1999). Appraisal theories. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion* (pp. 637–663). Chichester: Wiley.
Sloboda, J. A. (1991). Musical structure and emotional response: Some empirical findings. *Psychology of Music, 19*, 110–120.
Walker, B. N., Nance, A., & Lindsay, J. (2006). Spearcons: Speech-based earcons improve navigation performance in auditory menus. In: *Proceedings of the 12th international conference on auditory display, London, UK, June 20–23* (pp. 63–68).
Walker, B. N., & Nees, M. A. (2012). Theory of sonification. In T. Hermann, A. Hunt, & J. Neuhoff (Eds.), *The sonification handbook* (pp. 9–39). Berlin: Logos.

EMOTIONS, TECHNOLOGY, AND DESIGN

Edited by

SHARON Y. TETTEGAH
Professor, University of Nevada, Las Vegas
College of Education, Las Vegas, NV, USA

Beckman Institute for Advanced Science and Technology,
National Center for Supercomputing Applications, affiliate,
University of Illinois, Urbana, IL, USA

SAFIYA UMOJA NOBLE
Department of Information Studies, University of California,
Los Angeles (UCLA), Los Angeles, CA, USA

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
125 London Wall, London, EC2Y 5AS, UK
525 B Street, Suite 1800, San Diego, CA 92101–4495, USA
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, USA
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

© 2016 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-12-801872-9

For information on all Academic Press publications
visit our website at http://store.elsevier.com/

Publisher: Nikki Levy
Acquisition Editor: Emily Ekle
Editorial Project Manager: Timothy Bennett
Production Project Manager: Caroline Johnson
Designer: Matthew Limbert

Typeset by SPi Global, India

Printed and bound in the United States of America

ELSEVIER
Book Aid International
Working together
to grow libraries in
developing countries

www.elsevier.com • www.bookaid.org