The Cultural Politics of Television News Sound

A thesis submitted to the faculty of
San Francisco State University
In partial fulfillment of
The requirements for
The degree

Master of Arts
In
Broadcast & Electronic Communication Arts

by

James David Batcho

San Francisco, California

December 6, 2005

CERTIFICATION OF APPROVAL

I certify that I have read *The Cultural Politics of Television News Sound* by James David Batcho, and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirements for the degree: Master of Arts in Broadcast & Electronic Communication Arts at San Francisco State University

---

Dr. Chul Heo
Assistant Professor of
Broadcast & Electronic Communication Arts

Dr. Brent Malin
Assistant Professor of
Broadcast & Electronic Communication Arts

Dr. Vinay Shrivastava
Associate Professor of
Broadcast & Electronic Communication Arts

THE CULTURAL POLITICS OF TELEVISION NEWS SOUND

James David Batcho
San Francisco State University
2005

Sound in television news is dominated by voices of authority, resulting in a narrowing of ideas and a restriction of experiential and cultural representation. Through selective access and a removal of environmental ambient sound, television news operates noncontextually and thereby fails in its obligation to the public good. This masters thesis aims to shed light on the problems of sound exclusivity in television news, wherein historical approaches to methodology and microphone technology have resulted in practice that serves the needs of the broadcaster over the public. As part of this examination, I offer a re-evaluation of presidential candidate Howard Dean's infamous "scream" in Iowa and critique the practice of authoritative, subjective analysis in the broadcasting of the 2005 presidential inauguration. I argue for a more contextual sound practice through the inclusion of environmental ambient sound.

I certify that the Abstract is a correct representation of the content of this thesis.

_____   _____
Chair, Thesis Committee   Date

ACKNOWLEDGEMENTS

Thank you to my thesis committee—professors Chul Heo, Brent Malin, and Vinay Shrivastava—for the inspiration and guidance, for helping me to look at things in bold new ways, and for the friendship. I am grateful to my parents for always being positive and encouraging, and to Cynthia for the long conversations and her patience during challenging times. Barbara Teixeira and Adam Vadnais were my support group throughout the research process, and their input was invaluable. Finally, I wish to thank all the professors at San Francisco State University who have inspired me with their intelligence and wit.

TABLE OF CONTENTS

INTRODUCTION

Late in the evening on January 19, 2004, about 3,000 anxious people stood in the Val Air Ballroom in West Des Moines, Iowa eagerly awaiting the arrival of their hero. Affectionately dubbed "Deaniacs" for their youthful enthusiasm, these were the hard-core supporters of Vermont Gov. Howard Dean in his quest to become president of the United States. Earlier that evening, pollsters projected that U.S. Sen. John Kerry earned a surprise win in the Iowa Caucuses, relegating previous frontrunner Dean to a distant third-place finish. Until that point, the Deaniacs were riding a wave of momentum propelled by an unprecedented internet-financed campaign that led to overwhelming public support among Democrats. Their candidate led the polls leading into the caucuses, but things did not turn out as expected. The disappointing result was not lost on his supporters, who found themselves at a critical moment. This rally was an opportunity to show their candidate that their belief in him was still strong.

Sensing the urgency of the moment, Iowa Sen. Tom Harkin jumped on the stage prior to Dean's arrival and rallied the full house into riotous cheers and applause. By the time Dean arrived the crowd was fired up and the room was enveloped in a thunderous cacophony of sound. As supporter Doyle Fine II said in a guest column for the *Fort Worth Weekly*: "When Gov. Dean came on stage, the roar of the crowd was deafening" (2004). Dean picked up a wireless handheld microphone and spoke energetically through the ballroom's public address system about taking back the nation and about not giving up. The crowd intensity rose to meet his every amplified word, and as the people became louder, he responded with equal fervor. Eventually, the dynamic between Dean and his supporters rose to such a level, the crowd around him drowned out his voice. This was classic Howard Dean in his element. And to those in the crowd, he gave them all they hoped for.

Given the newsworthiness of the day's events—Dean's shocking loss and his anticipated public response—the media was also present in the Val Air Ballroom. Most of the major broadcast and cable television stations were on site to cover the event, hoping for live footage of Dean's arrival and speech. CNN's star political commentators, Paul Begala and Tucker Carlson, were there. So was ABC's Dianne Sawyer. The speech

was even broadcast live on C-SPAN and some cable networks. Watching the rally on television, the rest of America became witness to a very different occurrence from what took place in the ballroom (C-SPAN, 2004, video). Through the medium of television, Dean's speech exhibits a man seemingly out of control. His voice and energy level rising higher and higher, he seemed to shift from intense to borderline manic, finally reaching a crescendo of what could be described as a curiously over-enthusiastic scream. It was a yell so loud that it shocked the media and the television public. Over the next few days, the coverage was seen as so bizarre for a presidential candidate, it became a subject of continual repetition on news broadcasts. The event caught fire among the voting public and before long was christened the "I Have A Scream" speech. Dean became an object of ridicule in the press. According to the Center for Media and the Public Affairs: "Former Vermont Governor Howard Dean's coverage fell from 58 percent positive prior to Iowa to only 39 percent positive in the week between Iowa and New Hampshire primary" (2004). The candidate who once led the pool of Democratic Party candidates never recovered and quickly lost his bid to become the nominee. "The breathtaking media explosion turned the former Democratic presidential front-runner into a punch line and arguably hastened his campaign's free fall" (Associated Press, 2004).

To many in politics and news, the speech seemed to be the empirical evidence that proved their suspicions. That Saturday, weekend news commentators on major news broadcasts replayed the footage and used it to support a portrait of Dean as an unstable man who was unfit to run the country. But also, on that exact same day, a very different version of the event was launched on the Internet. This footage, videotaped from within the crowd by Dean supporter Joe Jensen, captured an entirely different moment (Jensen, 2004, video). From this view, we as viewers and listeners are witness to things more as they were seen and heard from within the event as it happened. Dean's voice could be heard in the context of the surrounding sounds of a throng of supporters whose voices rose like a swelling tide. As his voice lifted in intensity, the audience in attendance matched his enthusiasm. As he rifled off the list of states holding upcoming primaries, the people responded to each one with a deafening roar. Finally, when he reached the point of his infamous yell, the crowd was so loud it was barely audible through the supporter's camera microphone. The crowd had drowned out the man on stage. How could there be

such a difference between two versions of a single event? The answer lies in the practice of television news. In essence, Howard Dean's fate was determined by the particular practice of television sound.

THE VOICE OF TELEVISION

Discussions of television news aesthetics invariably concentrate on properties or metaphors of sight—among them, images, pictures, shots, framing, viewpoint, perspective, insight, and focus. These are all visually oriented terms. We are, after all, a visually oriented society attuned to the goal of orientation. It is the visual that strikes us with the most direct, cognitive sense of immediacy and helps us to make our world knowable. "[O]ur ability to retain precise visual information is far better than our ability to retain precise auditory information" (Alten, 1990, p. 558). Sound, in contrast, is something else entirely. It is generally taken for granted as an afterthought, or more particularly, as a problem to be contained. The reason for this is the presentation of the aesthetic.

Television sound, whether in practice or in academic discussion, overwhelmingly concentrates on a single source: the voice. Voices provide the much-needed informational aspect in televised content. The voice is also the audible component that we as human beings naturally have the most direct connection with. Our ears are tuned to its essential characteristics of pitch and frequency at the elimination of extraneous elements that may threaten its purity. "If a human voice is part of [a soundtrack], the ear is inevitably carried toward it, picking it out, and structuring the perception of the whole around it. The ear attempts to analyze the sound in order to extract meaning from it—as one peels and squeezes a fruit—and always tries to *localize* and if possible *identify* the voice." (Chion, 1999, p. 5) As such, sound, particularly in nonfictional television, is overwhelmingly *textual*. It is the written word, organized and pre-planned. Even in the coverage of live events, sound is either scripted ahead of time, or it is sculpted into familiar codes of language. This derives in large part from the cultural institution of television, with its well-worn modes of institutional and technological practices in which pre-conceived templates are called upon as a means to contain the unexpected. News broadcasts are designed to preserve and present a sense of finite "story" amid the ambiguity of an uncontainable world. "The most basic format consideration, which is shared by all news media, is that news becomes suitable only when it is transformed into a story" (Gans,

1979, p. 70). Voices offer the most manageable and direct method for structuring newsworthy *events* into news.[1] The picture element reifies these spoken scripts by presenting contextual imagery that provides the scene and background plus visual clarity and substance.

> No matter how live the image, the sound continues to serve as presentation, as commentary—in short, as audience to the image. Carrying out its italicizing function, the sound track serves a value-laden editing function, identifying better than the image itself the parts of the image that are sufficiently spectacular to merit closer attention on the part of the intermittent viewer. (Altman, 1986)

Television news therefore encodes a dual product for consumption, with sound and image reinforcing each other into a seemingly unified, broadcast-worthy story: The moving image elements provide the context, while voices offer story and structure through spoken words.

This means that in the production of sound, the primacy of the voice is preserved at all costs. Television voices are typically close-miked subjects or analysts who describe or debate the meanings of images. These are typically voices of some kind of real or perceived authority. Sound in practice makes every attempt to preserve this homogeneity, this singular sense of purity and predictability. Broadcasters present pictures as a neutral backdrop and thereby offer a presentation of objective, contextual realism against which the voice can be overlaid for structure. Sound is therefore not only textual, it is *noncontextual*. It is removed from the place in which an event occurs. Sound is both restrictive and additive rather than integral. Viewers are occasionally presented with so-called "natural" sound, but it is also tightly managed. During live events, its purpose is primarily to assist in the illusion of continuity across picture cuts. If it is mixed in later during post-production, it is re-constructed to suggest an *impression* of reality. In *live* news and event coverage, the overall practice of sound is even more cautious. Events as they occur in real time are unpredictable. But through rigorous modes of practice perfected through the course of the 20th Century, sound is not open but contained, and always in service to the demands of the news institution. Sound for television is a carefully constructed process that restricts access to the wider cultural and environmental character of an event and instead grants access to voices of authority that dictate meaning to an audience.

This naturally leads to the question of what gets left out. If television is dominated by selective voices, what is removed in the process? Sound designers typically refer to these externalities as ambient sound. Depending on the precise nature of this sound it is also called noise, environmental sound, or natural ("nat") sound. Ambient sound comprises three main categories: (1) the sounds of machinery, including electronic interference, "room tone," and the residue of sound technology itself; (2) the sounds of space, including direct-to-reflected sound and reverberation characteristics, and encompassing issues of intelligibility, points of audition, perspective, volume, and fidelity; and (3) the sounds indicative and representative of culture and society, including other, typically unmiked voices and the sounds inherent to a particular location. In all forms of classic media—broadly identified here as film, radio, and television—every one of the above categories of ambient sound constitutes an element to be suppressed, removed, or reorganized. (Reorganization is of considerable interest, for it relates to the practice of falsely reconstructing familiar *impressions* of reality, which carry clear ideological problems.) While all of these categories of ambient sound will be considered here, this analysis will concentrate primarily on the third category, and, where applicable, the second category. The sounds of machinery and the nature of spatial perception and representation have been well documented in recent qualitative academic analyses of sound, many of which will be considered here.[2] But surprisingly little has been addressed regarding the ambience of culture and the reasons for its removal, suppression, and/or reorganization in television practice.

Howard Dean's speech clearly indicates the need for such a discourse. In that situation, sound as typically practiced in television led to the downfall of a popular political candidate. The decontextualization of sound through the conscious stripping away of environmental and cultural ambience can be detrimental in the presentation of news events that affect not only the particular qualities and subtexts of an event, but the public response to such news reports. Why, then, is this the case? Why does television practice eliminate the environmental and cultural context that exists and is available in its natural environment? The answer is that efforts toward ensuring the primacy of the voice lead toward intelligibility in the voicing of language, and as an extension, meet the needs of the messenger—or more specifically, the broadcaster. In television news, sound as it

exists in nature is secondary to the intentions of the forces that present it for commercial interest. It is important to keep in mind that through the institution of television, sound is not simply a phenomenological issue. It is not simply a method of communication toward the understanding of an event. Such assumptions cast aside layers of subtexts that occur within every audible representation. Sound is rich with encoded intentions that often function to serve the corporate interests of the broadcaster or the subjective ideology of the speaker. While sold as a public good, television news is primarily a business, one vulnerable to the influence of hegemonic structures of control, capitalist motivations that trump the public good, and the resultant dumbing down of substance that is all too prevalent (Louw, 2001). The Dean example displays these tendencies very clearly. The Associated Press estimated that the "Scream" heard 'round the country was the result of 633 rebroadcasts of the pivotal sound byte (Associated Press, 2004). This is the result of an industry that spoon-feeds spectacle in an attempt to keep viewers tuned in to the broadcast. As a result, television sound is an issue of ethical concern more than one simply of aesthetic analysis.

This masters thesis begins with the premise that broadcast news is dominated by voices of authority and that sound practice functions to actively eliminate more pluralistic and open representations that may contextualize, and thereby challenge, these voices. I will attempt to examine the historical, institutional, economic, and cultural reasons why television news sound operates noncontextually, particularly in live event coverage. A theme of this examination is an analysis of the Howard Dean speech in Iowa. I will also offer a critique of live news event coverage through an analysis of the Presidential Inauguration of 2005. In the case of Howard Dean, two different versions of the same event emerged through two very different audible representations. Why did a more "open" version not receive significant airplay? More to the point, if a news broadcaster's stated aim of objectivity is genuine, why does it not represent sound in an open, contextual fashion?

I will attempt to answer these central questions and present a comprehensive critique of the practice of television news sound over three broad chapters. Chapter 1 attempts to answer the questions of how and why sound evolved into its current state in television. I will show how historical developments in aesthetic approach and

microphone technology resulted in a code of professional sound practice wherein selected voices dominate a soundtrack. Chapter 2 will explore sound as it is currently practiced in the broadcasting of television news by first examining sound theory and then moving into news gathering technique, "liveness," voice representation, and the commercial news industry itself. This chapter will also include a close textual analysis of the live broadcast of the 2005 Presidential Inauguration. Chapter 3 will look at the implications of current television news sound practices, examine the notion of news objectivity, and argue for a more contextual approach to sound representation in news. Throughout, I will reference Howard Dean's speech in Iowa—both the televised version and Jensen's independent footage—as an example of how the sound of events is represented by television.

CHAPTER 1. THE DECONTEXTUALIZATION OF SOUND:
AN HISTORICAL EXAMINATION

At the most fundamental level, Howard Dean's infamy was inscribed into the public consciousness through the use of two distinct but conjoined elements: a microphone and a medium. The historical progression of each provides clues as to how his Iowa speech and the public perception of that speech contributed to his eventual downfall. In examining the history leading up to television we see how two intertwined developments during the early years of film and radio led to the current aesthetic of television sound. The first of the two is technological: A competitive, capitalistic atmosphere in microphone research and development resulted in products designed to eliminate extraneous sound that might threaten the legitimacy of identifiable voices. The second is institutional: Television sound is borrowed from a model of radio and film aesthetics that favors intelligible information over the wider sonic environment of an event because such practice leads to predictable and manageable results for the industries themselves. As a result of these factors, sound practice today involves the active elimination of that which can provide wider cultural and representational perspective, that of ambient contextual sound. This chapter will focus on the technological and institutional mechanisms of film and radio during the 1920s and 1930s, a time when television was a mere experiment. It is during this formative age of audio reproduction and representation that sound became the medium it is today.

SECTION I — MICROPHONES: TECHNOLOGY OF EXCLUSION

"How sad this guy goes from 100 to 0 because of a good microphone."
- post by 'lia w. Shone' on valuejudgment.org, February 10, 2004. 02:08 PM

Sound as media began the moment efforts toward reproduction began. In 1876 Alexander Graham Bell invented the telephone and a year later Thomas Edison reproduced sound through the invention of the phonograph. Almost simultaneously, a sound transmission device and a recording medium came into being. As the two essential components of sound media—transmission and reproduction—both technologies turned out to be driving factors in the way sound developed in the early 20th Century. They also dictated the development of all other audio devices to follow. The first of these was the condenser microphone, which was conceived in 1881 (Altman, 1985), but did not become a practical device until much later. It was, in fact, more than four decades after Alexander Graham Bell earned his U.S. patent for his electrical transmission device (Eargle, 2001) that E.C. Wente of Bell Telephone Laboratories published a paper in *Physical Review* describing the electronic specifications for what constitutes the first condenser microphone (Wente, 1917). The work of Bell Labs was a sign of things to eventually come, with its engineers working on both telephony and microphone circuitry design. The basic sound transmission design was the same for both. "The microphone … amounts to little more than a highly sensitized telephone transmitter" (Pitkin & Marston, 1930, p. 221). While microphone design was about sonic realism, telephony filled a different need. The telephonic aspect was Bell's main business objective, and here, clarity in voice reproduction was a primary goal. "Intelligibility was clearly linked to conventionality at this early stage. Speech that could be easily interpreted on the basis of little actual audio information—a call, a query, a cliché—was more likely to be understood over the telephone's lines" (Sterne, 2003, p. 248). Bell Labs "felt entirely comfortable sacrificing 60 percent of the voice's acoustic energy (the lower frequencies) because they lost only 2 percent intelligibility in the bargain. The functional primacy of intelligible speech enabled telephone systems to reduce drastically the amount of power

required for transmission while retaining the ability to transmit voices with acceptable clarity" (Lastra, 2000, p. 164-165). Alongside Bell's practical compromising of fidelity to linguistic clarity was a more public, commercial goal, designed "to convince users that telephone conversation was the same thing as face-to-face conversation" (Sterne, 2003, p. 265). The point here is that Bell, a major innovator in microphone design, was primarily a telephone company whose engineers made transmission systems with the business goal of voice clarity in mind.

Of course, advancements in both telephony and sonic reproduction went hand in hand. In the late 19th Century, loose carbon granule elements combined with a moving armature receiver resulted in what can first be considered *the microphone* as used in telephony (Eargle, 2001). "The carbon granule transmitter and moving armature receiver complemented each other nicely; the limitations in bandwidth and dynamic range have never been a problem in telephony, and the rather high distortion generated in the combined systems actually improved speech intelligibility by emphasizing higher frequencies," (p. 5). The 1920s saw the emergence of new technology by Western Electric, the manufacturing arm of Bell Telephone, resulting in the electrostatic capacitor, or condenser, microphone (Eargle, 2001). One of the early technical issues to overcome in microphone design was in capturing the nuances of the voice, which in the early sound films was not captured with the sensitivity desired by studios. "In the first few sound pictures, very little difference could be detected between the voices of the men and those of the women" (Pitkin & Marston, 1930, p. 236).

Also during this time, a new radio industry was taking shape with the foundation of the Radio Corporation of America (RCA) in 1919 (Bruck, Grundy, and Joel, 1999). At this point in American history, capitalism had come to dominate as the mode of production, including labor, technology and the rise in new business opportunities (Kellner, 1990). The new sound reproduction technologies were in the process of migrating from an experimental form of communication and a military advantage into an emerging commercial broadcast industry. From its inception, radio sought to embody the aesthetic of controlled delivery. "Like the phonograph, radio technology was first conceived as a means of point-to-point communication" (Peters, 1999, p. 204). Originally, microphones of this early era were omnidirectional carbon mics (Altman,

1994) that captured a wide sound field. But in the 1920s, RCA engineers pushed to create more sensitive microphones, eventually leading to RCA's work on the ribbon microphone in 1925. In addition to the issue of sensitivity, radio studios needed a solution to the problem of sound reflections, wherein the architecture of the studio exhibited what was deemed unnatural reflections (British Broadcasting Corporation, 1951). Such reflections called too much attention to the *mechanism* of radio and compromised the goal of projecting a sense of intimacy and purity as delivered by the direct, unadulterated sound of the voice. Eventually, the bi-directional ribbon microphone became the preferred choice in radio. It offered enough desired sensitivity, isolation from room ambience, and directionality restricted to two on-mic sources.

The competitive environment of Bell Laboratories and RCA heated up in the late 1920s and leading into the thirties. During this time, developments in microphone design were expanding to meet the needs of the emerging film and broadcasting industries. Between 1931 and 1934, designs progressed at a rapid pace, leading to the development of the 77-series and constituting "the first unidirectional or 'cardioid' microphone" (Borwick, 1990, p. 11). Designed by RCA's Harry F. Olson, the 77-series was an "immediate success" (Altman, 1985) and marked a significant breakthrough in design because of its unidirectional polarity pattern. The design allowed microphones to narrow their sonic "focus," wherein sounds emanating from particular locations were isolated while peripheral sounds were rejected to varying degrees (based on distance and angle of approach). The model 77A set off a highly competitive research and development environment between Bell and RCA.

> While Bell engineers were scouring the engineering world for ways of reducing unwanted signals and enhancing the quality of desirable signals, at RCA Olson was looking in the obvious place for a solution to the same problems. Instead of assuring a high-quality signal by reducing distortion in the pick-up system (as Bell continued to do), why not design a method of aiming the microphone right at the desired sound source, thus eliminating numerous unwanted sounds? (Altman, 1985, p. 10)

This practice eventually led to more directional characteristics in microphone design, an engineering approach used in most on-location and radio microphones today. As the characterization suggests, unidirectional patterns reject external sound such as ambience, noise, reflections (reverberation), and any other unwanted sounds in favor of

those generating from a particular location—notably, *on-axis* or frontal sounds, and even more specifically, a particular voice. The very term "directional" was meant to indicate a direct-to-reverberant ratio, or a relationship between a direct sound and the reverberations that result from the direct sound (Eargle, 2001). The growing functionality of the unidirectional microphone marked a new opportunity in the radio broadcasting and film industries. Whereas an omnidirectional mic might capture sounds within a full 360-degree radial sound field, unidirectional polarity patterns allowed practitioners to isolate key points of emphasis *within* the sound field. Not only did these microphones emphasize particular sounds, they drastically reduced the so-called "noise" of the surrounding environment and the reverberant characteristics that affect clarity in the voice. "The desired high directivity results from the fact that only sound waves arriving on or near the axis produce additive pressures at the microphone diaphragm, whereas off-axis waves suffer varying phase-shift delays and tend to cancel. …[T]he high directivity action is restricted to middle and high frequencies." (Borwick, p. 107).

These mid-to-high frequencies are the area of the frequency spectrum where the human voice is at its most present and intelligible. The new designs could therefore not only isolate the voice, but also reproduce its most dominant characteristics. Also noteworthy is the way unidirectional designs enable a lengthening of the distance between subject and microphone. The *distance factor* of a microphone is a means of measuring how far the mic can be positioned from a subject before it represents the ambient pickup characteristics of an omnidirectional microphone. All microphones, regardless of their rejection pattern, will pick up more reverberant sound the further it is moved from a source. The more narrow the polarity pattern—and therefore the more narrow the angle of sound picked up—the greater the distance one could move the microphone from the subject before it took on a more omnidirectional behavior. A hypercardioid, for example, has a distance factor of 2; it can be placed twice the distance from a source as an omni before it takes on the more open direct-to-reverberant characteristics of the omni pattern (Eargle, 2001).

Examples of unidirectional microphone types fall into the first-order cardioid designs: the cardioid, the hypercardioid, and the line microphone. The first is the well-known heart-shaped pattern that exhibited a restrictive envelope that favored sounds

directly in front. The design was also meant to capture *off-axis* sounds, or those emanating from the sides to varying degrees, but there is a gradual increase in level attenuation the more off-axis the source of the sound. Off-axis rejection became more and more apparent in these types the further to the side and the more distant a source was from the front of the microphone. These mics also rejected sounds from directly behind the pickup unit, meaning that the microphone could be held in the hand without any so-called "handling noise." The cardioid handheld microphone is precisely what Howard Dean used in his speech in Iowa. Examining the C-SPAN footage, one can see that Dean is holding the microphone extremely close to his mouth, never more than roughly two inches away. This positioning leads to a highly directional sound envelope that would achieve excellent ambient cancellation characteristics. Also, as the crowd becomes louder and louder, Dean's voice increases in volume to adjust, which further cancels sounds outside the envelope. This is clearly evident in videographer Joe Jensen's footage from the crowd (2004).

Hypercardioid mics serve a different function. Directionality is increased because the distance from mic to source is typically longer. This is a characteristic that became useful on film shoots because the distance from source to microphone is longer than that in a radio booth, for example. Hypercardioid designs offer "the highest directivity obtainable from a first-order pressure gradient microphone" (Borwick, 1990, p. 105). They use a parabolic reflector to "ensure that all sounds arriving in line with the axis are bounced toward the microphone, giving a considerable gain in axial signal level…" (1990, p. 106). Sound waves arriving off-axis are "scattered and do not contribute much to the microphone output." Line microphones, also known as shotgun or rifle mics, result in even greater off-axis rejection. "As a general rule, the line section of the microphone is added to a standard hypercardioid microphone, and it acts to increase the DI [directivity index] of the hypercardioid above a frequency inversely proportional to the length of the line" (Eargle, p. 110). A variation on the cardioid family is the Figure-8, a bidirectional design that favors sounds coming from two opposite directions along a 180-degree axis. These mics became useful in radio broadcasting for two-subject voice reproduction, such as in-studio dramas and interviews.

Microphone development in the early 20th Century provided the technological means toward isolating and thereby more clearly reproducing particular signals in a sound field. As suggested by the work of Bell Labs, and followed soon after by RCA, the voice was invariably the favored signal worth preserving. This competitive developmental atmosphere offered steady enhancements toward the preservation of clarity in voice reproduction while excluding undesirable sounds inherent to a particular location. Such advancements helped Bell's and RCA's economic prospects because they met the needs of the emerging broadcast and film industries. These industries required predictability in voice reproduction in order to maintain control over particular messages, a trend that has become commonplace in today's television news media.

SECTION II — AN EMERGING PRACTICE:
THE INSTITUTIONS OF RADIO AND FILM

"I wish I could convey the atmosphere... the background of this... fantastic scene."
– Carl Phillips (fictional character) in Orson Welles' October 31, 1938 broadcast of
War of the Worlds

Sound in television today is a rigorous and intransigent code of expectation. Method is little more than a template, one so ingrained that it is almost a pure science today. Sound engineers are regarded for their ability to solve problems that may interfere with the recognized codes of predictability. In order to understand how these precision standards of practice evolved to their current state in television, it is necessary to look at the history of sound as applied to the business of content creation. The story of sound's evolution is one of competition in both economics and philosophy. It is an interweaving narrative of technology, innovation, capitalist interests, and competing ideas on the meaning of representation and the preferred approach to sound practice within such arguments. This last issue is a particularly complex dynamic, one involving battles between sound engineers and their colleagues and employers over everything from ontology to the nature of reality to peer respectability. Along the way, all of these interests helped to build a practice that today goes unquestioned.

Sound, as applied to the cinema, was in turn transferred from classic theater (Auslander, 1999). From its very inception, sound in film did not so much fill an artistic or informative need as it did a attention generating one. James Lastra, Jonathan Sterne, Rick Altman and other academics of film sound history have chronicled how sound in media began as an endeavor of purely capitalist intent. Theater owners of the early 1900s would hire live performers as a means of luring potential patrons to the stage. "Until the early 'teens, sound worked to solicit patrons and draw their attention—a function that was to support a more general strategy for directing spectators to performers in the theater, or to objects in the image. In fact, the direct address made possible by sound would shape most silent film sound in one way or another" (Lastra, 2000, p. 99). Lastra

documents how early sound was used "as a means of attracting attention to the Nickelodeon itself" (p. 98). Various sound-generating schemes such as trap drummers, piano players, and Edison's phonograph would draw the attention of passers by on the street toward what the theater had to offer on that particular day. For Lastra, this method of applying sound to visible objects constituted an early form of the "direct address" method of sound practice that is so prevalent in television today.

Eventually, these sidewalk performers became part of the early silent Nickelodeon films themselves, bringing sound to what is inaccurately regarded as the "silent" film era.[3] Drummers in particular were so adept at making a wide variety of noises that they can be considered the earliest "sound designers." Altman specifically points to the year 1909 which began the "overtly discursive vaudeville practice of using sound effects primarily for comic effect" (1994, p. 15). These early sound performers were "effects specialists (typically drummers working a series of 'traps') who labor mightily to reproduce every possible sound suggested by the image." Playing along to the silent film during playback, these artists became worthy performers in their own right.

> By adhering to a practice that was discursive and performative, the effects man *created* new hierarchies within the image, drawing spectator attention to incidental features because they could make noise. Under his gaze, the image ceased to signify in a predictable way, but became a pretext for virtuoso displays of sound. (Lastra, 2000, p. 105)

Sound became so prominent in this regard that it threatened the traditional superiority of the image. Aside from the live performance aspect, sound offered something the image could not—a direct and immediate sense of connection with the audience that was distinct from the more evocative nature of the image.

> [Sound] addressed audiences directly and offered the structural conditions for shaping audience response… [W]hile the represented space of the film became more and more distinct from the space of the audience, sound retained and continued to develop a mode of sonic 'attractions' maintaining a direct relationship of performer/audience copresence. (2000, p. 106)

Clearly, sound was not integral to the image in the "silent" film era, but abstracted from it. The traditional visuality of filmmaking until this time was not only a form of entertainment, but also a recognized craft that stood alone with a sense of artistic integrity. In contrast, the additive component of sound developed as a side show,

something to sell tickets. "Discursive, participatory, emphatically present, and lucrative for exhibitors, sound constantly threatened to usurp the image" (2000, p. 102). During this silent era, when sound was separated from the image, the idea of synchronizing sound to film was already becoming a business even though the process had yet to be tested (Kellogg, 1955). Proponents of synchronous sound were not even sure how it would be used. Edward W. Kellogg notes that using it for speech was seen as dubious, given that pantomime and titling "had been so highly developed, that giving the actors voices seemed hardly necessary, although readily possible" (1955). However, given the capability to make actors speak, the transition to the synchronous sound film was indeed marked by the use of synchronous dialogue to great effect.

> When in 1927 such a picture was shown *(The Jazz Singer)* the story, the music and the dialogue were splendidly adapted to produce a fascinating picture with great emotional appeal, in which no element could have been spared without serious loss. In short, the excellence of showmanship played no small part in making it clear to everyone who saw it that the day of 'Talkies' was here. (Kellogg, 1955, p. 356)

Such enthusiasm was not matched by filmmakers of the time who had spent decades perfecting their silent, visually oriented craft. "[A]ny form of sound accompaniment that threatened to disrupt the priorities of image space by redirecting attention to narratively inessential elements similarly threatened the norms of the hierarchical and linerized images" (Lastra, 2000, p. 110). Many filmmakers justifiably saw studio intervention as heralding an era of bad art, resulting from difficulty adapting to the new mode of practice (Kellogg, 1955). Adding to the disdain directed toward the talkie was the proclivity on the part of studios to inject singing wherever possible (Pitkin & Marston, 1930), simply because the popularity of certain songs at the time presented a marketing opportunity. A hierarchical audiovisual relationship developed that not only placed images above sound on the level of perceived artistic value, but also clearly defined their aesthetic roles. Pictures were regarded as the foundation, while sound served a very different purpose—a method "to address the audience directly—to hail them." (Lastra, 2000, p. 120).

> As the image-based narrative began to assert its dominance … sound's functions began to change. Sound could serve as a clarifying addition, as in the lecture, or it could provide important narrative motivation for on-screen events by supplying a missing sound cue. … On the other hand, music and commentary could work against narrative unity—by punning, by meta-commentary on the situations or

characters. (2000, p. 118)

This distinct separation in approach between audio and visuals helps to understand why sound is practiced differently from image making and why it serves a more practical than artistic need. This direct address, hailing aspect is prominent in radio and would later become the principal aesthetic of nonfictional television.

As sync sound films increased in prevalence, new codes began to emerge which served similar capitalistic purposes. The meta-commentary Lastra refers to was presented most substantively through voice content. The increasing prominence of the voice within the soundtrack gave Hollywood a means of commenting on the visual component. As a result, cinema could deliver authenticity through perceived realism. "[S]ound apparently offered another type of symbolic capital that, itself, could be exploited to enhance revenues: realism" (2000, p. 106). Mike Wayne and other theorists of the sound-image relationship refer to this clarification through encoded reality as "verisimilitude," which refers to the manner in which "sound is used to make the image 'credible' within a very narrow definition of 'realism'" (1997, p. 176).

> The conventions of dominant cinema require the soundtrack to be integrated with and subordinate to the image track, which is privileged as the dominant carrier of meanings. The task of sound in this conception is to support the 'reality' or more accurately, enhance the verisimilitude of the image and narrative. (p. 178)

Once sound became a regular component of production in film, struggles emerged over its ontological signification. The initial instinct was for the microphone to maintain the same position as the camera. In the silent era, film cameras maintained more or less a static shot length for duration of a scene. Performers would change their position relative to the camera through blocking and movement. "In the old silent picture, the director was absolutely bound to the camera. He had to hold a shot up to its end, and then title it. Hence, the entire movement of the picture was restricted to whatever the photographer could record" (Pitkin & Marston, 1930, p. 206). In the early days of location recording, the microphone would move with the camera to its eventual shooting position and thereby take on the same perspective as the resulting shot length. Through tighter framing, close up shots would have a tighter sound "focus," resulting in clear and present dialogue with little environmental sound or reflected sound as a result. This matched how

a person in an audience might experience the act of listening to someone in close proximity to him or her on a personal level. Conversely, a wide shot would exhibit more reflected or ambient sound—the natural sound of the surrounding environment and the spatial properties of a place—and less direct sound of the voice itself.

When editing became more commonplace, a sound problem immediately ensued that led to one of the great early debates of the 1930s, and continues to this day: What happens to perceived sonic space during a cut between a long shot and a close-up, or vice-versa? Walter Murch perhaps best describes the phenomenon of the cut in film at the visual level as the action of the eye blinking (2001). When we change our attention in our daily experience from viewing a wide landscape vista to a foregrounded object within it, we typically blink as a means of reorienting ourselves to the new focus of perception. This is why, according to Murch, we can accept the shift that comes with the cut in film. By contrast, sound somehow offers a different state of perceptual awareness, wherein a long shot begs for less direct sound with a close-up needing more direct sound. This cutting from a distance-oriented sound field to a close-up sound field resulted in two problems for studios: 1) such shifts risked losing the clarity in sound—something studios were concerned about, given the uneven quality of sound reproduction at the time (Pitkin & Marston, 1930); 2) such sonic shifts would call more attention to the cut itself, threatening studio producers' desire to keep the mechanism of film invisible. "Here verisimilitude is in conflict with narrative intelligibility. … Extended use [of cutting] throughout a conversation risks drawing the audience's attention to the constructed nature of the sound/image relationship and therefore risks the rupture of verisimilitude" (Wayne, 1997, p. 183-184). If there is a scene involving two people talking within a large crowd, the intelligibility of the dialogue will dominate the mix even though this is not the way one might hear it naturally.

> Two equivalent processes are at work in the sound perspective debate. In one, the sound film posits a single, original, unified audio-visual event that precedes its representation (here, sound and image perspective roughly coincide). In the other, the disjunction between sound and image perspective is 'healed' through the very *discursive* nature of a hierarchical sound recording, mixing, and reproducing practice that privileges intelligibility of the voice. (Wurtzler, 1992, p. 98)

These problems of perspective and continuity resulted in disagreements between

sound practitioners and their employers, each with differing opinions as to how sound should be represented in film. Technicians favored a "general model of perceptual simulation." which placed importance on "…absolute, perceptual fidelity as the obviously paradigmatic objective" (Lastra, 2000, p. 138). Hollywood, by contrast, deemed that this model based on simulation of acoustic space for a central auditor failed to account for the more pressing need for sound to be predictable and clearly understood.

> This [fidelity] model ran headlong into the competing demand for narrational clarity. In fact, many of the transition-era problems seem to have revolved around two competing and more or less incompatible models of sonic space and intelligibility—ones tied to different forms of sound practice and different representational norms, and which therefore supported different representational industries. (Lastra, 2000, p. 138)

As a result, two schools of sound philosophy emerged that were in stark contrast: a *phonographic* or "fidelity" model, and a *telephonic* or "intelligibility" model (2000). "[T]he former sets as its goal the perfectly faithful reproduction of a spatiotemporally specific musical performance (as if heard from the best seat in the house); the latter, like writing, intelligibility or legibility at the expense of material specificity, if necessary" (p. 139). While Hollywood saw both models in practice during the 1930s, the telephonic approach was most widely evident, particularly in the case of voice representation.

> According to the principle of narrative priority, dialogue recording tends almost uniformly from the early thirties on toward the telephonic, minimizing the amount of reverberation, background noise, and speech idiosyncrasy, while simultaneously maximizing the "directness" or "frontality" of recording, and the intelligibility of dialogue. (Lastra, 2000, p. 139)

Conversely, the phonographic model of sound lobbied for the encapsulation of a wider sonic space. In practice this involved, first, several microphones capturing an omnidirectional presence, and later, a single-mic approach where the mic assumes the same position as the camera (Altman, 1992b). Since the camera embodied the perspective of the spectator, a single mic would act as audible perspective, thus capturing a natural sound space. The closer a person was to the camera, the higher the levels of direct sound captured; the further from the camera, the more reverberant the sound (higher reflected to direct ratio). In other words, camera and microphone cohabitated the same perceptual location. The microphone simulated not only camera position, but also auditor position.

This approach "fixed the position of the microphone as an audience member and allowed all the nuances of the space and performance to color the resulting recording" (Lastra, 2000, p. 146). As the performers move positionally in relation to the camera, the sound would reflect such changes in position through a change in direct to reflected sound.

While the phonogenic method of sound found its way into film, it did not become preferred practice. The reason is the new capabilities of microphone design and how they enabled the studio producers to solve the problems of cut discontinuity and clarity in voice recording. The *boom microphone* was a breakthrough in this regard when it became available in 1929 (Lastra, 2000). Here, a unidirectional condenser mic is placed on a long pole that was carried on set by a sound technician. The boom mic follows the action and results in continuously close-miked situations, generating high intelligibility representations regardless of shot length. "With this new system … the microphone is perpetually kept within approximately the same distance of the speaker, thus canceling out nearly all the factors which the earlier system retained" (Altman, 1992b, p. 53). This *close-miking* approach marked a significant shift in practice toward the intelligibility or telephonic model as the prevailing mode of practice that is still used today. It's a practice that preserves the quality of the voice at the expense of ambient, environmental sounds, and the more random elements in production that producers found threatening. The boom mic could also move *through* space and single out particular sounds whereas the phonogenic approach to mic setup simply *captured* space. "The advantage of the boom-mounted microphone on location, as in the studio, is that by varying its distance to the sound source you can make it reflect the focal length of shots." (Alten, 1990, p. 499).

This shift was not simply driven by the technology, however, more that the technology enabled the studio to achieve its aims. Clear speech was paramount for the studios, and the practice of intelligibility provided it. "[T]he reproduction of speech was implicitly (and often explicitly) guided by the preordained role of speech as, above all, an *intelligible* conveyor of information" (Lastra, 2000, p. 145). Those with the power in Hollywood were now able to dictate methodology to the sound technicians. Altman notes that the practice of preserving clarity over realism became something for which sound technicians were "praised and rewarded" by the studio (Altman, 1992b, p. 54). By 1938, the intuition of sound engineers moved toward intelligibility over spatial accuracy

(Lastra, 2000). The question of why sound in audiovisual media evolved the way it did can be answered in pure economics. The studios, enabled by technological advancements that eliminated undesirable sound, were able to provide a more predictable, manageable, and therefore more salable product by preserving the homogeneity of the voice. As a result, studios, not sound artists or film directors, set the template for an aesthetic approach to sound that remains relatively unchanged.

Another result of telephonic practice was a uniformity in film soundtracks. The desire for predictability in sound, and clarity in voices in particular, led to a dogged pursuit of consistency in volume levels. Where sound engineers fought for a simulation of acoustic space—including variations in dynamic range to match physical distances in relation to the camera, and nuances in direct-to-reflected sound—those in positions of power found the solutions they needed to common problems through a non-dynamic soundtrack. "The construction of a uniform-level sound track, eschewing any attempt at matching sound scale to image scale, thus takes its place alongside the thirties' numerous invisible image-editing devices within the overall strategy of hiding the apparatus itself, thus separating the spectator from the reality of the representation situation..." (Altman, 1992b, p. 61). This is indicative of the studio effort toward transparency in sound—not transparency in the sense of openness, but as a means of projecting an impression of invisibility in the encoding process. As with editing, the desire through sound was to remove any impression of the *constructive* aspect of any audiovisual text. As Jonathan Sterne points out, this is a continual effort in sound reproduction—a strong desire to liberate any notion of the apparatus of sound from the eventual representation. As a byproduct of this effort to remove threats to intelligibility, something else is lost as well—the human externalities inherent in the production itself. "Wishing away the noise of the machine then suggests wishing away the noise of society. The relations and functions that made possible the moment of sound reproduction were labeled *exterior*, outside the act itself" (2003, p. 259).

The combination of technology plus a desire for intelligibility marked the end of the fidelic model of representation as a preferred method of practice. "[C]onflicts over representing sound space were settled de facto by the adoption of the standard of close miking and a certain 'frontality' of address. … [C]lose miking is relatively contextless or

spaceless… [S]uch sound is spatiotemporally *unspecific*, legible beyond the original context... more like writing than speech" (Lastra, 2000, p. 148-149). Some filmmakers actively rejected the strict practice of intelligibility in their work. Rouben Mamoulian, for example, found ambient sound to be an important representation of cultural identification. Lucy Fischer, in her article *Applause*, analyzes the soundtrack of the 1929 Rouben Mamoulian sound film of the same name (1985, p. 240): "The undecipherable background dialogue, the vaudeville music, the street noises, all construct for us a sense of the material world which the characters inhabit." Fischer says that Mamoulian went against standardized sound practice in order to better represent the environment he was filming. "For it was the aim of most dialogue engineers to eliminate all ambient noise in order to foreground the spoken text. Thus technicians sought to counter the omnidirectional quality of early microphones, which tended to pick up all background sounds within their range. But for Mamoulian, ironically, these noises were crucial, and were to be accorded equal status to speech" (p. 239-240).

Such approaches were rare, however, as mainstream Hollywood embraced the politics of exclusion. Over the course of film's development, such treatments became sonic signifiers of the works of avant-garde auteurs working well outside the mainstream. Jean-Luc Godard, for example, reinvigorated an approach to realism in film by abandoning Hollywood's constructivist aesthetic as an effort to "interrupt any kind of ideological illusions" on the part of the audience (Shrivastava, 1996). By the time Godard entered the world of film in the 1950s, cinema already had its sound aesthetic firmly in place. At that point, television was the new audiovisual medium. And as a medium of pure intelligibility also borrowed from yet another industry, radio, it fully embraced the telephonic approach to sound representation. Through film, television had its model, a mode of practice that fused the voice driven radio aesthetic with the intelligibility approach to film sound.

THE INFLUENCE OF RADIO

In 1921, the first commercial AM broadcast from KDKA in Pittsburgh, PA was launched (Bruck, Grundy, and Joel, 1999), and two years later, network radio was born. Powered by the Radio Corporation of America, the industry quickly expanded. By the time *The Jazz Singer* was projected in cinemas in 1927, the exclusively audible medium of radio was already firmly established. While film had to negotiate the contrasting philosophies of sound positioning in relation to picture and continuity issues, radio had no such audiovisual conflicts. As a purely sonic medium, the practice of radio was one that strived for discrete, isolated clarity in voices and sound effects. Radio was also predominantly studio-based compared to the more location-based medium of film. What the two shared was a desire to remove any notion of apparatus from the product. In contrast to film, the problem for radio was much simpler—removing the studio from the voice. Omni-directional mics captured excess room tone and reflections, which called too much attention to the artificial, manufactured presence of radio and not enough to the more *telepathic* intentions of radio, which were geared toward generating an illusion of interiority. Radio's early history was an effort toward regaining a sense of direct communication that was lost through a broadcaster's abstraction from an audience. "New forms of authenticity, intimacy, and touch not based on immediate physical presence had to be found" (Peters, 1999, p. 214). The technological solution to this problem was the use of bi-directional ribbon microphones (Altman, 1994) that were able to isolate a pair of voices and remove unwanted reflections.

> Since a microphone of this type responds less and less as the direction of the sound departs from normal [on-axis], it picks up much less reverberation (random in direction) than a nondirectional microphone having the same sensitivity for sound of normal incidence. (Kellogg, 1955)

As the sound films grew in popularity and prosperity, the film industry needed practitioners. Naturally, it looked to radio. "For where had Hollywood found its sound technicians? By far the majority… had come from the radio studios. The early years of sound cinema were thus heavily marked by the version of reality offered by other modes of representation—first silent cinema, then radio" (Altman, 1992b, p. 55). These radio

engineers had already developed the ability to isolate and discriminate among particular voices and relational sound effects. When they migrated to film, they applied their highly developed, institutionalized template of interiority to a medium that was one of visual exteriority. The radio aesthetic brought an audible template that was simply applied to the film image.

Altman expands on the outcome of this shift in his article *Deep Focus Sound*, which in part addresses the development of what he calls radio's foreground/background aesthetic that reveals a striking similarity to the intelligibility model of film sound. "Two regions dominate: the foreground, in which actors move and narratives develop, and the background, which serves to guarantee diegetic reality while concealing discursive reality (by reducing the camera's ability to register any space not identified with the diegesis)" (Altman, 1994, p. 9). This reflected a desire for the enhancement of primary speech and hard effects while placing background effects and "unimportant speech" in the distance. Radio was a practice wherein every element had its diegetic position. These discrete positional arrangements were not only determined by volume and presence in the mix; they arose from institutional compartmentalization. "To understand radio of this period, we must thus recognize in each diegetic volume variation a potentially overdetermined discursive variation, in each narrative character an actor with discursive responsibilities, in each diegetic crescendo a sound with a potentially discursive function" (p. 8). The foreground/background aesthetic of radio paved the way for the intelligibility model of sync film, which by 1938 had become the dominant form in cinema (Lastra, 2000).

It was also in 1938 that Orson Welles broadcast *War of the Worlds*. It created an impression so seemingly realistic in delivery that it shocked many listeners into believing that Martians had invaded Earth. The sensation of his radio play led to film offers from Hollywood. In 1941, he released *Citizen Kane*, widely regarded as one of the most influential films of all time. The movie was groundbreaking in its use of sound, including overlapping dialogue, meticulous sound perspective, and layered sound effects—all borrowed from radio practice. "Intelligible dialogue remained central throughout this period" (Altman, 1994, p. 8). It is within this context of new forms of industry, technology, and practice that *Citizen Kane* was released. "*Citizen Kane* constitutes the first effective meeting place of the century's two most powerful media, broadcasting and

film. As such, it is in a sense the first modern movie, the first film to structure itself around the discursive necessities of a broadcast medium" (Altman, 1994, p. 25). As Altman observed, *Citizen Kane* is regarded for its realistic approach to sound though its "debt to radio realism" (1994, p. 1). But it is not radio's realism that is represented in the film, rather the institutional practices that marked the *business* of radio at the time. "Certainly, the primary auditory novelty of *Citizen Kane* lies not in its sound realism, but in its intermittent adherence to radio's model of narrative/discourse interplay" (p. 24). Radio is less a medium of realism and more a commercial industry that uses sound to manage its requirements of predictability. Through *Citizen Kane*, Welles offered more to the future of television than to the film industry, according to Altman. "If we seek the progeny of *Citizen Kane*'s radio sound, strangely enough, it is to television that we must look, for that is where we eventually find the short-unit rhythm, differing volume ranges, and discursive/narrative struggle characteristic of radio and *Kane*" (p. 27).

  Citizen Kane reveals for Altman "six distinct reality codes" (p. 23), among them, separate but similar codes for cinema and radio:

> No. 5. "**cinema**: clear separation of sound into foreground and background zones; foreground sound must be clear and spaceless; background sound may lack clarity and express space, but must be reduced in volume to assure comprehensibility of foreground sound; only point-of-audition sound regularly links the two zones and treatments;" (p. 23)
>
> No. 6. "**radio**: discursive sounds are always loud and comprehensible; either on- or off-mic, narrative voices and effects are either spaceless and comprehensible or distanced and narratively unimportant; transitional sounds constitute the major exceptions to this rule." (p. 23)

It is here, in this convergence of cinema and radio during the early forties, that we find the model for television sound. In both, sound is relegated to narrational clarity in voice recording, a hierarchical structuring of necessary and unnecessary sounds (even unnecessary voices), and a disregard for the spatial characteristics and representational potential of ambient sound. Through cinema sound practice, television had its audiovisual model of continuity and vocal clarity; through radio sound, television had its interiority and hierarchical selectivity. What is lost in such an approach to sound representation is a

the broader sonic environment of an event. As such, television fails in its enormous potential to present wider inclusivity through sound. Often, as exemplified by the coverage of the Dean rally in Iowa, the ambient sound is what reflects the larger social interaction occurring in the event. ABC News' Diane Sawyer admitted as such, saying that the foregrounded, single-microphone audio stream used for the Dean rally did not give an accurate portrayal of the environment from within the hall in which it took place. "[W]hat we heard on the air was not a reflection of the way it sounded in the room" (Sawyer, 2004). "In the room, the so-called scream couldn't really be heard at all." With the room relegated to background, to something that may usurp the conventions of recognized practice, the audience is incapable of receiving the larger significance inherent in the event itself.

CHAPTER 2. SOUND IN TELEVISION

When television came into being in the late 1940s it was a mirror image of the already established radio industry. Not only was radio's sound practice in evidence, but the network business model, the same programming choices, even the same high-profile on-air talent migrated from radio to television (Kellner, 1990). Television continued radio's interior sound aesthetic, and the rise of news broadcasting expanded the need for a voice-dominated approach to sound that continues today. It is interesting to note that in roughly 60 years of commercial television, its approach to sound has undergone no substantial changes in news broadcasting. Throughout that time, film has come to embrace the layers of artistic and representational potential that ambient and nondiegetic sound can provide, and the music recording industry has dramatically increased in advanced approaches to engineering and production as the result of widespread experimentation. Television's only significant alteration was the Federal Communication Commission's shift from monaural to stereophonic sound in 1984, nearly thirty years after the music industry had already made the shift.[4] Even with the change, the monophonic mindset remained in the production of television news reporting. The same goal is at work: Production practice dictates a preservation of the voice at the expense of all other sound elements in order to not threaten the perceived legitimacy of the broadcaster. There is a sense of faith that the picture element of television will provide the reality against which the voice can be overlaid almost like a transparency.

In fact, "transparency" is a good metaphor for considering the role of sound in television. With live news or otherwise reality-based content, we have at our disposal a pair of moving layers, alive with activity. Imagine we lay these atop an overhead projector that allows us to see both of these layers presented on a screen. The first layer is highly visual and complex, dense and multidimensional. It is filled with pictures, actions, color, depth, dimensionality and a wealth of context to help us orient ourselves to the activity it presents. Laid upon this foundation is the second layer, this one completely translucent. Written on this transparency are clearly legible words in thick black ink.

Nothing else exists on this layer. Nothing is projected other than the black words and the translucent invisibility of the transparency itself.

The first layer, the base layer with all that visual context, is video; the second, the translucent one with words scrawled in black marker, is audio. Together, they comprise the audiovisual marriage of television news. Continuing with the metaphor, the audio layer is distinct in that the text is written by handpicked subjects—usually persons of some kind of authority or prominence. What this second overlay does not show is *context*. Other than the written words, this layer is invisible. All of the context we find with the visual is intentionally removed from the audible to maintain the purity and integrity of the text. So where pictures display a richness in experiential representation, sound is the equivalent of black text laid on top of the visuals. This means that sound—manufactured and highly subjective—is always vulnerable to the more ideological mechanisms of media.

The point behind this metaphor, is very simple: pictures show and words describe. That word "show" is of course a visual word. We have no sound equivalent. Television rarely "shows" sound in the same way pictures are shown. This is an important distinction because it means that the picture element of media is *provided*. Granted, it is framed and therefore determined by human activity. But compared with sound, it is offered to people, presented in such a way as to give members of an audience the ability to judge for themselves the meaning of what they witness. The visual, while framed and composed, still manages to provide a less adulterated, less manipulated form of contextuality. Sound, on the other hand, is not provided; it is dictated through analysis, description, and narration. It is noncontextual.

SECTION I — TELEVISION SOUND THEORY

There is surprisingly little qualitative research on television sound in a cultural or theoretical sense. On one level, this is attributable to the still prevalent status of sound as inferior to the more thoroughly studied visual element. Sound is not only monophonic in its presentation, but also simplistic in the way it is regarded. As has been mentioned previously, it is a problem to be solved or a need to be fulfilled rather than a representation of a given situation. Perhaps as a result of this, practitioners of sound have long been regarded as technicians more than artists. As an extension, writings on sound tend toward the analytical or, more specifically, the methodological and technical. There is a wealth of popular sources that offer instruction on how to effectively produce sound for television, but little on its cultural significance. For this reason, we have to look at writings on its predecessors—film, as the aesthetic medium of influence, and radio, as the practical and institutional medium of influence.

One of the most problematic aspects of sound is the relationship between reality and its more discursive attributes. Christian Metz once famously wrote that a gunshot heard in a movie theater is indistinguishable from a gunshot heard in the real world (1985). This statement has been almost universally derided by sound scholars for several reasons, among them: (a) systems of representation (in particular, microphones) are directional and therefore the spatial properties of sound change depending on the perspective from which it is captured; (b) every sound carries encoded meanings that go beyond simple acoustics and into areas of initial intent, ideology, and economics. A third reason arises from the first two assumptions: Any sound in media is never an original; it is by its very nature, a copy. "Recording did not simply capture reality as it was; it aimed to capture reality suitable for reproduction" (Sterne, 2003, p. 236). In this regard, sound is a very different medium than visuals. Not only is it a historically additive component (a trend that continues today), it is a manufactured component that serves a particular purpose, driven by whatever intent is desired. This makes sound an enormously complex aesthetic element, rich with subtexts and codes used to achieve particular ends.

31

Perhaps the strongest and most prevalent subtext for this discourse is the additive nature of sound, particularly in film. Throughout its historical relationship with the cinema, sound has been used to reify the legitimacy of the image through a process of restricting sound in production and reconstructing it in post-production. Stephen Handzo illustrates the degree to which environmental sound is manufactured in film practice for manipulative intent. "[T]he ambient (background) noise, although *intrusive,* is essential to *convey* reality" (1985, p. 384, emphasis added). Mary Ann Doane describes how sound is synchronized to the image in postproduction to create an illusion on behalf of the viewer that there is an unquestionable unity at work. "Sound itself is often described as adding life to the picture. And the life which sound gives is presented as one of natural and uncodified flow" (1985a, p. 57). She adds that any visible subject, for example an individual, is a body that "must be anchored in a given space." Lastra notes a similar phenomenon at work. "The most important aspects of a sound were assumed to be its origin, its identity, and its comprehensibility" (2000, p. 138). Sound is therefore a means of upholding an illusion of audiovisual unity to stave off any cognizance on behalf of the viewer that film is actually a constructed series of discontinuous edits.

Sound offers to uphold this illusion through manufactured continuity. "The continuation of the same sound over a cut on the image track diverts attention from that cut" (Altman, 1992b, p. 57). In live television events, ambient sound is strictly controlled, managed, and more often than not, as in the case of Howard Dean, eliminated almost completely. For recorded events intended for later editing prior to broadcast, ambient sound is employed primarily for constructed realism and to facilitate continuity in editing, for much the same reason as in film. During production, sound technicians will record a period of "room tone" or "nat sound" in order to later mask continuity problems across edits that are compiled in post-production (Weis, 1995). "Ambience from the production track is generally considered important primarily for creating a seamless sonic landscape, helping to create the illusion that all the cuts of the scene are taking place in a continuum" (Sonnenschein, 2001). In film, this is done for artistic reasons, where illusion is part of the medium. But this constructed illusion exposes the central problem of sound in televised news: The event is compromised by the need to create continuity through artificial representations of sound. There is an irony very much at work here: In

television's desire to project an appearance of objectivity, it deliberately *manufactures* the one audible agent that may in many cases offer such contextualism in the first place. This is simply another example of the approach to sound in television—as a tool for manipulation rather than a legitimate entity itself.

Ambient sound is not the only reconstructed element in post-production—the voice is fused to images after the fact. The typical news report approach is to show the anchorperson or location reporter on camera, then cut to video of something he or she is describing. This footage is known as "cutaways" or "B-roll" (Compesi, 2000). More often than not, no sound of the event itself is audible unless it serves to simply legitimize a sense of perceived realism in the coverage of the event. Instead, the picture is deemed sufficient for context and the now off-camera voice is used *to describe the picture*. This is simply a redundancy of the image with nothing of the sound that exists within the diegesis of the event. This, again, is the hierarchical nature of audiovisual media. The picture element always dominates and sound is used to uphold the legitimacy of the image. Within this typical example, however, there is another hierarchy at play, this one within the audio mix itself. On nearly every occasion, the voice is the dominant audible form, pushing contextual sound aside. "The stratification, the continual subdivision which the soundtrack undergoes, is aligned with the aim of sustaining a rigid hierarchy of sounds. …[A]bove all, [it is] the intelligibility of the dialogue which is at stake…" (Doane, 1985a, p. 58). Doane sees "the practice of using speech as a support for the individual…" (p. 58). Michel Chion agrees, saying that in film "there are not *all the sounds including the human voice. There are voices, and then everything else.* In other words, in every audio mix, the presence of a human voice instantly sets up a hierarchy of perception" (1999, p. 5).

Speech itself is also stratified into layers of subcategory. Herbert Zettl identifies three forms of speech in audiovisual media—dialogue, direct address, and narration (1999). Narration and direct address are the categories of speech that dominate television news as a means of gaining control over a message or a series of moving images. Chion calls such voiceover content "textual speech" (1994, p. 172). Textual speech has the power take control over narrative, setting, and moment, and in so doing, achieve domination over the audiovisual presentation. "If textual speech can control a film's

narration, of course, there *no longer remains an autonomous audiovisual scene*, no notion whatever of spatial and temporal continuity. The images and realistic sounds are at its mercy" (p. 172, emphasis added). Again, Chion's concentration is film, and in a film context, such dominance is temporary and reserved for key moments of emphasis before being released back into the diegesis. But in aesthetic terms, the same form of textual speech is delivered in television news. The difference is that whereas film's use is temporary, television's use is continuous. Voices of description and analysis are unceasing. This is evident in every aspect of televised sound, from reporters on the scene, to anchorpersons, to the analysis of specialists, to the carefully crafted sound byte. As such, broadcast news is continually usurping the sonic character of the event in favor of controlled ownership over the event. This is a manifestation of the need for power on behalf of the broadcaster. "Textual speech is inseparable from an archaic power: the pure and original pleasure of transforming the world through language, and of ruling over one's creation by naming it" (Chion, 1994, p. 173). As a result, the environmental, ambient character of time and place as it happens is never allowed to emerge with any clarity through the nonstop textuality of television news sound.

The externalities of this process of continuous textual speech is the ambient sound of the event being covered—an element either eliminated or manipulated for reification of the spoken word. What is interesting about ambience is that while it can be used for ideological purposes through artificial reconstruction in post-production, it is also the means by which we may achieve audible context. While ambient sound *can* be used for ideological purposes, if used more responsibly it is also the one sonic element that reflects the wider environmental or cultural atmosphere of a situation. This is the great tragedy of the radio aesthetic in televised events: the voice stands in for that which tells the story on its own. Chion—who defines ambience as "sound that envelops a scene and inhabits its space, without raising the question of the identification or visual embodiment of its source" (1994, p. 75)—offers an interesting way of thinking about ambient sound. He refers to ambience as "territory sounds, because they serve to identify a particular locale through their pervasive and continuous presence." What is even more significant in Chion's view of such sound fields is that he approaches ambience not as a representation of place but a means of *determining* place. "In the case of ambient sounds … what is

important is the space inhabited and *defined by* the sound, more than its multisource origin" (p. 79, emphasis added). In other words, such sound defines the location in which events occur. As an extension, ambient sound is also a means of offering a wider sense of cultural inclusion within a representation. The environmental context that is included through such methods widens the range of ideas and in turn, broadens the discourse of an event.

Returning again to Howard Dean's speech in Iowa, we can now gain a better understanding of what happened that night. Again, there were two video recordings of the same actual event. First was the C-SPAN audio feed, which offered only one sonic element—a single, unidirectional microphone, fed monophonically through a sound board, and broadcast through cable affiliates and into people's homes. This is recognized practice in news audio feeds. In every sense, the process adhered to solid, highly developed practices of professional sound production for television. The microphone choice was perfectly suited to keep the sound of the environment (including reverberation and crowd noise) from overwhelming the legibility of the central figure speaking to a crowd; the public address system ensured a clean signal flow; Dean's technique of microphone-to-mouth proximity was just right to maintain clarity in his voice; and the home viewer could hear every word he was saying.

By contrast, there was independent videographer Joe Jensen's footage from within the audience of 3,000 supporters (Jensen, 2004). Here is a completely different representation, one characterized by high levels of reverberant-to-directional sound, an increasing loss in intelligibility over time, and generally bad practice for television coverage of a man speaking on stage. But it is here that we also find a prime example of the fidelic approach to sound that was dismissed by the late 1930s. The difference is that the contextual ambient sound, the environment of the place, provides the answer to why Dean's voice elevated so high in volume and why he screamed into the microphone. The fidelic representation exhibits very clearly how within the context of the raucous throng of supporters, Dean's voice became unintelligible *to the people within the event itself*, independent of the broadcast. Without the benefit of a monitor feed from the camera or broadcast studio, he had no way of hearing the way his voice was coming through the broadcast medium. The only audible source of reference he had were the loudspeakers

coming over the public address system, something the crowd had almost completely overwhelmed. To further understand the implications of these two very different representational approaches, we need to move beyond the theoretical aspects of sound and examine the technical and commercial methodology toward audio production in live news events today.

SECTION II — THE PRACTICE OF TELEVISION NEWS SOUND

"Now I'm just going to be quiet and let you listen."
- Bernard Shaw, CNN, reporting from Baghdad, 1991

Any book on sound methodology for television will instruct that in news field production, the voice should dominate and that ambient sound should be removed through microphone choice and placement (Compesi, 2000). "Microphones affect ambience, frequency response, loudness, signal-to-noise ratio, and texture; in short, they affect the overall quality of recorded or transmitted sound" (Alten, 1990, p. 301). Audio capturing is a highly technical process of maintaining control over a sonic environment. "The main challenge in doing a production on location is to record the principal sound source with little or no sonic leakage from either the acoustics or other sound sources. Therefore, you must judge a recording site on your ability to neutralize unwanted sound" (Alten, 1990, p. 288). The BBC's guidelines for location recording dictate that in noisy locations "it is almost essential to use microphones with special discrimination against sound from remote sources" (1951, p. 14). If the event is intended for live use, unidirectional microphones are used to strip away all but the voice.

In situations such as Dean's, it is common practice and therefore not surprising that the live public address engineer would offer him a unidirectional handheld. What was missing, however, was wider coverage of the event through the placement of additional microphones, something that is rarely done. Electronic field production (EFP) and electronic news gathering (ENG) teams typically have far less time for setup compared to the meticulously structured shots associated with cinema. Stanley Alten perhaps best sums up the sense of urgency that is commonplace in sound production captured on location: "In news the immediacy is more important than the aesthetics, but this is not to say that news producers ignore production values. They simply do the best they can under the circumstances" (1990, p. 465).

It is true that in production, legibility in voice content is important for presenting the text of a situation or event. Good production involves the filtering of hum, rumble and

interference that may interfere with the main sound element and disturb the listener. "Noise—unwanted sound—is enemy number one in audio production" (Alten, 1990, p. 36). Microphone choice facilitates this process of elimination. The preferred microphone for broadcast news is a unidirectional, cardioid or hypercardioid design, one that captures sound from only one location, namely, the voice. Another common approach is to use a lavaliere microphone, which is typically attached to a subject's clothing. "[T]he advantage of this technique is that dialogue is clear and present, with a minimum of background sound. The disadvantage is that sonic perspective, regardless of a shot's focal length is the same" (Alten, 1990, p. 501). All other sounds are therefore rejected (Compesi, 2000). Unfortunately, an extension of this practice is that production sound becomes an effort in stripping away referential sound that threatens to remove the context of the event being mediated. Zettl advises that "all major moving and stationary sound sources must be covered with microphones" (1999, p. 309). However, this is rarely practiced in live television and the re-broadcasting of live events because it is often deemed unnecessary in relation to the voice. This type of "coverage" was not evidenced in the Howard Dean footage. Again, we cannot ignore the fact that in television event coverage, sound is more of a problem to be solved than it is an art form. If a sound technician feels that he or she is clearly hearing the words of the subject holding the microphone, his or her work is done. This is what general practice dictates. Location audio for television constitutes a series of problems, and the major problem gets solved first. Invariably, this is clarity of the voice. An additional component of the Dean situation is the combination of the event and the coverage of the event. Dean's voice is being generated both for the Val Air Ballroom audience, through a public address (PA) system, and for the television viewer through the broadcast. The voice goes first to the PA and from there to an auxiliary output fed to videotape. It is easier to solve the main problem of direct voice broadcast than dealing with the many issues of public address and its inherent acoustic properties. John Eargle says that the difficulties with sound reinforcement in such situations include reverberation time, direct-to-reverberant speech level, and signal-to-noise issues (2001). "[A]t some given distance, the reverberant field begins to dominate. The reverberant field is fairly uniform throughout the room, and at distances far from the loudspeaker, the reverberant field is dominant" (Eargle, p. 216). A

television sound engineer would know this in advance and probably not want to deal with the difficulties, particularly since the feed from the live sound board would achieve the principal goal. So there are two distinct sound engineering tasks at work here: (a) the hired public address engineer providing clear voice amplification for those in the room, and (b) the broadcast engineer offering clear voice reproduction for the broadcast designed for the home viewer.

This need to strip all artifacts of the mediation process from the informational message is a strong one, and cuts to the heart of a great deal of philosophical thinking on communication. In his introduction to *Speaking Into The Air* John Durham Peters compares this practice to a Heideggerian concept of telepathy (1999). Heidegger's argument postulates that communication is at its purest when it mimics the "intermental plumbing" of the mind (p. 18) and attempts to become a transparent conduit of ego to ego. "The dream of perfect communication through semantics recapitulates the dream of telepathy, a meeting of minds that would leave no remainder." Peters takes issue with the fantasy of telepathy and, in the process, hits on a central issue—that in television's effort to provide a unadulterated construction of message, what is lost is the nuance of experience.

> Again, communication as bridge always means an abyss is somewhere near. And even in the propaganda view, the antinomy recurs: communication working telepathically has an evil twin in the specter of the mesmerized mass in the clutches of the leader, just as the fear of closed consciousness appears at the mass level as the fear of the lonely crowd, atomized and mutually oblivious. (Peters, 1999)

At issue is fragmented identification, the loss of pluralistic involvement within the space of events. The *textual* message becomes more important than the *contextual* message of the people, atmosphere and cultural or environmental presence that make up the wholeness of the event. Zettl makes matters worse when he advises that the voice take prominence over the event. "The information function of sound is to communicate specific information verbally. In our verbally oriented society, a word is often worth a thousand pictures" (1999, p. 314). What Zettl fails to recognize is the inherent subjectivity of the voice that works as a mediator of reality and therefore does a disservice to the representation of events as they happen. This is an ongoing ethical

problem in mediated communication, whose nature is one of authoritative distance from the context in which events occur. Decontextualization, therefore, is essentially the removal of anything that might offer a challenge to the dominant voice of authority. As such, the practice of sound involves the removal of that which may offer a broader sense of understanding for the listener. The danger as it relates to television is that such an effort toward exclusion through audio practice creates a separation between the event as it occurred and the event as presented. Granted, sound once recorded is never a pure reflection of reality because it is done from a particular perspective through microphone coverage. But that doesn't mean that better representations can't be enacted through human practice if the goal of context is considered as paramount.

Interestingly, we can look to independent media as the example that puts the notion of "sonic reality" into broader perspective. Ever since inexpensive camcorders became widely available to the public, individually videotaped events have presented a counter to the mainstream media. Often, these documents capture a contrasting notion of an event in ways the major media companies can't capture because of the restrictive nature of their standardized practices. Camcorders are able to produce a more open representation of a sonic environment than the expensive production rigs used by ENG (electronic news gathering) teams. As heard in videographer Joe Jensen's footage, a camcorder often provides a more open sound envelope because of the less directional and more inclusive polarity patters in these microphones. A notable subplot is that Jensen very quickly made his video available to the mainstream media so they could show it over their commercial broadcasts. The broadcasting of such "amateur" video often receives airtime when such footage reveals a new twist on an existing story. Jensen's footage, however, received very little attention. The question must be asked: Why did television continue to show the close-miked version of the Dean speech even though there was additional coverage that put the event into context? The answer may lie in the notion that more open forms of representation would have reduced the spectacle of the event. And in a capitalist-driven media, the spectacle is what generates interest and, as a result, economic benefits. There is also the issue of repeatability. Without spectacle, repeatability might present an appearance of bias in the news organization. Tim Graham of the conservative *National Review Online* had already criticized ABC for questioning

its own coverage of the event, claiming that the network was more concerned with "complaints about damaging an ultraliberal candidate with an anger-management problem than it is about presenting an appearance of objectivity" (Graham, 2004). If ABC had repeated Jensen's version from the crowd as much as it repeated the close-miked version—there would have been accusations that ABC was forcing its own agenda as a means of dissipating what was already decreed in the public consciousness as the already recognized truth of the event, which appeared to be a confirmation of everyone's fears: an unstable candidate.

THE IDEOLOGY OF TELEVISION NEWS SOUND

It is often said that television is a reality-based medium. In cinema, the film itself is the text. Television, of course, develops its own series of texts, but it is not a closed system. Television news promises to present a reflection of the outside world. We can argue the level to which the medium reflects that world, but it does carry a debt to the public. Television news engenders a sense of *responsibility*—not to the medium, but the event and the people interacting within that event. Somewhere within this event lie nuances of experience that should be represented by the broadcaster. There are many factors involved regarding sound as heard through the medium of television, some of which have been discussed here: the restrictiveness of microphones, the need for selectable components, the aesthetic treatment of the sound, the "grain" of the recording device itself. As Rick Altman accurately notes in *The Material Heterogeneity of Recorded Sound*: "Recorded sound … always carries some record of the recording process, superimposed on the sound itself" (Altman, 1992a, p. 26). Altman, however, goes too far in his view that by virtue of sound's existence as *representation* rather than *reproduction*, sound therefore necessarily "creates an *illusion* of presence" (p. 29, emphasis added). While it is true that sound is *affected by* the multi-tentacled apparatus of the recorder and its operator, it is a stretch to assume that all representations thereby constitute pure illusion.

A more tempered view of this is what Chion calls "mediated acoustical reality" (1994), or what Philip Auslander calls conditioning through "mediatization" (1999). The conditioning of sounds toward particular purposes becomes a foundation upon which *notions* of truth emerge. Through a listener's ongoing familiarity with the media-enhanced voice—one that is made prominent through flattering microphones, equalization, and compression—the ears will naturally gravitate toward such conventions and find legitimacy in what is spoken. "For natural voices are not only produced but also heard by unconscious comparison with mediated ones, which tend to stand out more" (Chion, 1994, p. 104). In other words, the institutional practice itself is what encodes the authority into the voice that is speaking. The difference between Chion's and Altman's

views is that from Chion's perspective, sound is not so much inherently misrepresented. Rather, a hierarchy is constructed among the various voices that may exist that favors the one that the broadcaster deems as worthy of prominence. It is another manifestation of the interiority of radio. Auslander refers to people at sporting events who bring radios or small televisions because it is the particular quality of structured sound and prominent voices that lend an impression of reality to the event. In cases such as these, the audience is "trying to achieve a kind of aural intimacy that can be obtained only from the *reproduction* of sound" (1999, p. 35, emphasis added).

Auslander goes further to suggest an interesting aesthetic reversal that occurs based on the conditioning of an audience. The live can now only take on authority when it has been mediatized. "[W]hereas mediatized performance derives its authority from its reference to the live or the real, the live now derives its authority from its reference to the mediatized, which derives its authority from its reference to the live, etc." (Auslander, 1999, p. 39) And "...once live performance succumbs to mediatization, it loses its ontological integrity" (p. 42). While Auslander is talking about all media—sound and image—we can certainly find within this example the authority that the voice brings to the television viewer. Meditization of the voice through microphone design and mixing approach creates its own aesthetic of manufactured attention and its own form of authority by artificially extracting us from the sounds in the natural world. Television becomes not a cite of the real, but a separation from our possible connection with presence.

Steve Wurtzler examined this phenomenon by categorizing the relationship between spatial and temporal presence (1992). Howard Dean's speech would fall into one of two of Wurtzler's positions, depending on whether someone witnessed the speech in person or on TV. Those in the room at the time of the speech would fall into Position I— temporal simultaneity and spatial co-presence. Those watching at home on television would bear witness to a different relationship with the event, Position II—temporal simultaneity and spatial *absence*. Position I involves being "wholly present" in the moment while Position II constitutes "simultaneous presence and absence, a combination of qualities of both the live and the recorded, the immediate and the *mediated*" (1992, p. 90, emphasis added). This helps to further explain why it is impossible to reverse what

happened with the coverage of Howard Dean in Iowa. No matter how much Dianne Sawyer may report on the inaccuracy of the coverage based on discussions with various cable and network executives,[5] and no matter how much she says after the fact that what the American people saw was a *mis*representation, we can't accept such an abstraction. We view the scream as we originally heard it because television listeners have learned to accept the microphone as the audible agent of authority. To a television viewer, whatever comes through a vocal microphone is reality. It is only when we look at the footage shot by Joe Jensen—a more "natural," less mediatized representation of the sound of the event—that something seems inauthentic in the *television broadcast*. Here we are separated from our standards of normalcy and orientation. In this footage, all we find is low-fidelity sound, with no audible anchor to hang onto and no authority figure to guide us through. In a sense, Jensen's footage—with its ambient sound and sense of chaos— brings us closer to the spatial co-presence found in Wurtzler's Position I. As a result, Jensen's footage becomes a subject of fascination, perhaps even a version seemingly closer to reality, but not one of prominence and authenticity that comes with the mediatized version. Returning from his footage to the C-SPAN coverage, we as viewers still find the latter humorous and unusual, because, again, that is our sense of *televised reality*. Even with the knowledge of the other point of audition, we connect more substantively with the mediatization.

To get a better understanding of this idea of mediatized reality, we again turn to film sound as a progenitor to television. In this case, it was Hollywood's division of reality into interior and exterior categories, a practice undoubtedly influenced by the radio aesthetic when it was applied to sync film.

> In the arguments over sound perspective, 'realism' (as an effect of the ideology of the visible) is viewed as conflicting with intelligibility. If the demands of sound perspective are respected (that is, close-up sound 'matches' close-up picture, long-shot sound 'matches' long-shot picture), at a certain apparent camera-subject distance intelligibility of dialogue is lost. (Doane, 1985a, p. 59)

If there is a scene involving two people talking within a large crowd, the intelligibility of the dialogue will dominate the mix even though this is not the way one might hear it naturally. Doane says that this is Hollywood's way of rationalizing its own definition of realism, which is split into "psychological" or "interior" realism and "visible" or

"exterior" realism. "The truth of the individual, of the *interior* realm of the individual (a truth which is most readily spoken and heard), is the truth validated by the coming of sound" (1985a, p. 59).

Television, as a medium of individual talking heads, takes this idea of interior realism to extremes. It has adopted only one half of the duality of Hollywood realism and made it the anchor of truth in all sound representations. There is an internal sense of justification that so long as the individuals with microphones are clearly heard it is presenting an accurate story. The other half of Hollywood realism, that of the visible/exterior variety, is never actually real at all by any definition. Instead—if it is used at all—it is to uphold the validity of the psychological/interior "realism" of miked voices. As mentioned previously, ambient sound, or the non-close-miked voices of a crowd, is used as an ideological construct so long as it can be controlled. This tool is manipulated to provide an illusion of context if such manipulation serves to lend credibility to the voices speaking or to the images being presented. It is the same form of processing that is used to enhance the entertainment value of particular live programming such as sporting events. Here, microphones meticulously cover the scene because it serves the entertainment value of the program. "The sound of the crowd is essential to the excitement of most sports events." (Alten, 1990, p. 478). Alten says that while opinions vary as to how to approach live sporting events, crowd noise is an effective tool toward the entertainment value of the program. "Burying the announcers' voices in the crowd sound to make the announcers part of the event and to unify overall intensity," (p. 480). This shows the degree of control the mechanism, by way of the institution and the technician, have over content in ways that far surpass anything available regarding the visual. It is clear how this power over audible representation can be used toward ideological ends when transferred to a news context.

THE VOICE OF AUTHORITY

As for the voice itself, it is important to address what is gained by the broadcasters and lost to the public through the isolation of particular voices in live news reporting. It should not be assumed that in a critique of the restrictive nature of voice dominated content it is also by extension a criticism of discourse. Rather that in the isolation and magnification of particular voices, the cultural environment of experience is discarded. In the coverage of live reports or events in the field, as is so common in news, the environment is where the living, breathing, active energy of human activity resides. In many situations it is within this cultural environment that we as viewers and hearers can make determinations on events in ways that an individual voice of the network may either misrepresent or not fully appreciate. A journalist or chosen analyst cannot possibly understand every nuance that is taking place in a given situation, particularly when it is presented through Westernized training and practice. And yet, it is in these voices that the institution of journalism pins its reputation. It is therefore worth asking the question: In eliminating the ambient context of the surrounding environment, what is society provided through isolation of the voice?

Essentially, a television audience is granted access to various manifestations of authority. Those provided with microphones fall into several categories: the *subject*, the *specialist*, the *correspondent*, the *commentator*, and the *anchorperson*. All those allowed to speak to the public are either employed by or invited to participate by the broadcasting institution. The *subject* can take on many forms, from the President of the United States down to the man on the street. But he or she is first asked (or occasionally cajoled) to participate by the broadcaster. With the occasional exception of the "everyman" type *subject*, all voices bring with them some presence of perceived authority. "Professional codes ensure that what is considered important is that which is said and done by important people. And important people are people in power" (Kellner, 1990, p. 113). Consider the typical television news interview. In economic stories it is often a business leader; in a strike it is the head of the labor union; in a story of war it is the daily news briefings of politicians. We may see *pictures* of all of these scenarios as context—a shop

keeper sweeping the floor, a lock on a chain link fence outside a factory; a group of people surrounding a crashed helicopter—but we never hear such context. Any extensive analysis or perspective is conducted at a distance through studio commentary by the network itself or by those deemed worthy of the network. The *specialist* on the other hand is typically one of a panel of experts who debate specific issues. "The debaters are either government officials or representatives of recognized institutions, and as such they inhabit the sphere of newsworthiness" (Carpignano, P., Andersen, R., Aronowitz, S. & DiFazio, W., 1990, p. 115).

Kellner sees these authority figures as unwitting servants who uphold the broadcaster's need for moderation in an effort to communicate to the public that all is well. "[T]elevision news attempts to mediate between the opposing factions on different issues, policies, and ideologies and to promote a middle-of-the-road consensus, flattening out differences and managing conflicts" (1990, p. 115). What is rarely represented are views that reside on the fringes of political or social thought. Granting access to such figures may threaten the appearance of controlled objectivity that the news media relies upon. "Television news usually reinforces existing opinions; it is not a forum for new ideas or critical perceptions" (p. 114). *Anchorpersons* and *correspondents* offer a different form of authority, in this case the charming presence of reason and authenticity. "A good anchor is a good actor and with the lift of an eyebrow or with studied seriousness of visage, he or she can convince you that you are seeing the real thing, that is, a concerned, solid journalist" (Postman & Powers, 1992, p. 31).

To understand the degree of removal *from* events that these authority figures exhibit, we can borrow again from film sound and consider the interplay of diegetic and nondiegetic sound. In film theory, *diegetic sound* refers to the sound of the space within a given story (Doane, 1985b). Examples of diegetic sound are character dialogue, ambient sound, sound effects, and a song playing on a radio in a character's home. *Nondiegetic sound*, by contrast, consists of any sound separated from the story space and whose purpose is to add narrative or emotional content strictly for the audience's benefit. Examples of nondiegetic sound are a musical score and voiceover narration. We can apply the diegetic/nondiegetic relationship to television in the following way: The space of events, the diegesis, is exhibited by ambient sounds, the interactions of the public

within a particular location, environmental sounds, and the specific sounds indicative of a particular cultural setting. These sounds fall into the diegetic category because they occur within the spatial and temporal location of the event as it happens. Closely miked voices of authority, however, constitute nondiegetic sounds in the same way a voiceover or music soundtrack is heard in film. These nondiegetic voices are abstracted from the event as it unfolds. These voices therefore have the ability to dictate notions of meaning upon the event as it occurs in much the same way a film's narration injects significance into a narrative scene.[6]

If practiced with respect to the integrity of the event itself, beyond the ideological needs of the messenger, the voice can offer a great deal to the understanding of the event. But far more often, the voice becomes a means of dominance *over* the ideas. Michel Chion says that in film "a human voice structures the sonic space that contains it" (1999). In television, however, the voice becomes a force that *overwhelms* the sonic space that is separated *from* it. Even more problematic, the space is where the event actually occurs, not in the human analysis of it. Imagine any situation in which there is live news coverage of an event. Think of what might happen when microphones isolate voices and leave the surrounding atmosphere excluded. Imagine a world leader who gained power in a military coup, speaking of the next great frontier for his country while a woman in the crowd, unheard, cries out for justice against tyranny. Imagine any situation in which some voice of authority is telling the world how things are, while the surrounding sounds, excluded in the broadcast, tell a completely different story. Imagine again the situation of Howard Dean speaking to his most enthusiastic supporters that night in Iowa. Dean was the star and "the story," but he was not the event. The event was the public, silenced by the apparatus, which *made* the event. Their enthusiasm was the reason the words were spoken through the microphone as loudly as they were. Their enthusiasm, and their participation, was never represented by the broadcast news industry.

SECTION III — THE LIVE EVENT

On semantic grounds alone, the concept of the *live televised event* is an interesting notion.[7] There are two components at work here. There is the *live*, what Zettl calls the "open future" of the world at large fused with television's technological ability to capture and broadcast such timelessness (1999). Concurrently, there is the *televised event*—the shaping of moments as they occur through images and sound. Television always attempts to manage both, and in so doing, to project an impression of controlled significance. In this sense, the broadcasting of the live event is television's version of improvisational theater. Much of the appeal of live television is that, as with improv, we never as viewers know quite what might take place. We are witness to the uncertainty of interaction— perhaps something dramatic will happen, perhaps someone will fail miserably. Maybe it will give us, the viewing public, something to talk about the next day.

Howard Dean is the embodiment of this, and an example of why we are fascinated by the live event. As television viewers we live in the eternal state of possibility and even expectation that in the live moment, something like The Scream will become the moment that people talk about the next day. Television as a business depends on such human curiosity. There are several forms of live events that television carries. They can be completely unplanned captured moments, something organized but whose form is unknown ahead of time (such as the Dean situation), or they can be ceremonial events wherein a general script is known ahead of time and television provides the coverage. In this last category, a type of live coverage that has not been discussed here thus far, the news analyst or commentator takes center stage. This is where the newsman as analyst mediates between event and viewer, someone who stands "between us and them, between private and public tensions" (Campbell, 1987, p. 337). This is what Daniel Dyan and Elihu Katz call the "coronation" (1992).

> Television not only confers roles upon principals and viewers, but serves as simultaneous commentator and subtitler of the event. The more remote the viewer from the event—physically and psychologically—the more the broadcasters help to 'bring home' its meaning and enliven that meaning with interest and relevance. (1992, p. 38)

This is television in its most self-congratulatory state, when the interior voice of knowledge and empathy scrawls its message upon the external world of action. Through the trustworthy newsman, the script of life is voiced in reassuring phrases and as a result, meaning is provided for the viewer.

> In the case of Coronations, television rehearses the audience in the ceremony they are about to witness, carefully spelling out the meaning of the symbols, framing the event by separating it from daily life, monumentalizing it, upholding its official definition, and offering a story line and commentary to shape its interpretations. (p. 38)

Presidential inaugurations are prime examples of the coronation. They are presented as the quintessential live televised ceremony, in which the broadcaster guides the public through a process of American political significance. A network's star anchorperson leads the show and is supplemented by celebrity commentary carefully chosen by the network. The 2005 inauguration in particular is notable because the previous 2001 inauguration of then president-elect George W. Bush resulted in very visible displays of protest. These protests dominated the coverage—constituting what Dyan and Katz would call a "hijacked event" (1992)—and cut through the guarded ceremonial style typical of such coverage. In 2001, the unplanned elements threatened the sanctity of the coronation and *became* the story. With knowledge of this precedent, I videotaped the 2005 inauguration of George W. Bush for his second presidential term. During the procession to the White House, I isolated the central 20-minute section where the president's motorcade passed by citizens gathered in the crowd, including the "sanctioned" area designated for protestors. During the taping, I used a remote control to conduct a live scan among a few of the major news broadcasts—specifically, ABC News, CNN, Fox News, and MSNBC—to examine the coverage of the event in real time and tape the process.

- **00:00 — Starting at CNN.** Wolf Blitzer (off-camera) is in the studio talking to a reporter (also off-camera) standing near an area of the inaugural procession that the government cordoned off, or "sanctioned," for protestors against President Bush. Blitzer: "All right, Bob, I'm going to interrupt you. We're having a little trouble hearing you. It's very noisy where you are." He then offers commentary on the protestors' interaction with police that he just removed from the broadcast.

Blitzer: "They're [police] becoming a little bit more agitated as the noise goes right in front of them."

- **00:50 — Switch to ABC News.** The motorcade continues through the location of the sanctioned protestors, and an amplified voice of some unknown origin becomes clearly audible within the *diegesis* of the event. Peter Jennings of ABC notes this (speaking to field correspondent Terry Moran, both men off-camera): "Terry, it sounds to me like one or more people along the route have the capacity to have their voice amplified." Moran replies that someone is using an amplification device to "try and get both the president and the media to hear what they have to say." Moran assumes that the amplified voice belongs to one of the protestors. Jennings: "We don't really know on television, watching it on the monitor, whether or not the sounds we hear are from the people who are somehow related to the picture." He then introduces George Stephanopoulos and the two continue to discuss the ceremony.

- **03:06 — Switch to CNN.** Analyst Jeff Greenfield (off-camera) is in the process of describing the state of mind of people who are in political opposition to the president's policies. Live video footage shows brief pictures of protestors with signs depicting Bush in the likeness of Adolph Hitler. Greenfield speaks of the concept of opposition, not in relation to civilian protestors but rather elected Democratic leaders: "You know, one of the things about politics—I used to do it for a living before I became pure and a journalist—is that when you lose, it hurts." He continues his analysis and concludes by saying: "So, sure, it would be unimaginable that supporters didn't outnumber protesters here by a huge margin." Greenfield is making a mistaken connection between civilian protest and the Democratic Party legislators.

- **04:51 — Switch to ABC.** At this point, Jennings and Barbara Walters (also off-camera) are ending a discussion. Then Jennings: "At the moment, based on what we're hearing, as well as we're seeing, there are a lot of people out there. We've seen bigger crowds on Pennsylvania Avenue, but this looks to be a pretty big crowd. It's a little hard to tell if it's friendly or not because the *sound* [he emphasizes that word] is largely that of people who are clearly antagonistic to the

president or to his policies." During Jennings' analysis, the previous amplified voice is heard again. Instantly, Jennings stops talking in mid-sentence in order to try and listen to the voice. This is the result of the conditioning of journalists, whereby the amplified and therefore *constructed* voice is the one worthy of attention. Within the crowd of many, someone has suddenly emerged with a particular mediatized clarity that, while not perfectly legible, has the aesthetic of authority. Inside the fidelic space of the event, a voice of enhanced intelligibility has emerged in the diegesis. This is one of the rare moments in any of the broadcasts where there is a pause in close-miked analysis and description. Before long, Jennings resumes speaking: "I can't understand what we're saying." (He means "…*they're* saying," but given the meditization of the voice, his slip is notable given the level of control his network typically has over the sound.) Jennings then asks Moran in the field: "Terry do you understand what they're saying?" Terry Moran explains that he has learned that the man on the bullhorn is an official of the ceremony and not a protestor. He says that the amplified voice is directed at Bush, "…emphasizing to the president—I don't know if he can hear him inside the limousine—all the people who have come out to wish him well." Jennings, seemingly embarrassed by his earlier assumption, says: "Okay, so that's a cheerleader somewhere along the way." Jennings then turns to another of his celebrity analysts—NPR and BET news interviewer and commentator Tavis Smily (also off-camera)—to have a discussion about the topic of protests. The two speak mostly in clichéd sound bytes. Smily: "We live in a country where protest has its place. It's protest that makes this country so great." He describes how the media tends to have a "rather condescending view of protests, whatever the party in power might be." Jennings replies with a laugh: "Tavis, if you think we are condescending to the demonstrators, goodness I hope you'll say so." While Jennings and Smily are having this discussion, the ABC footage cuts to an area separated from the motorcade, where cameras have captured an incident in which protestors are shaking a chain-link fence dividing them and a throng of police officers. The protestors can be seen kicking down sections of the fence and throwing snowballs at the riot police who respond with tear gas. Roughly 20

seconds of footage passes by before the camera zooms out and pans to show visually where this is occurring in context to the motorcade. There is no sound of the goings-on. Instead, what you hear are Jennings and Smily having a good-natured laugh about their profession. Jennings makes no mention of the incident and again returns to discussing the man with the amplified voice. This situation verges on Dyan & Katz's notion of a "hijacked event," one in which the broadcaster must decide whether to "be faithful to the agreed definition of the event as a ceremony" or to shift to a "breaking news" mentality (1992, p. 72). ABC in this situation chose to maintain the ceremony of the event through the continued banter between Jennings and Smily.

- **08:16 — Switch very briefly to CNN.** No coverage in pictures, words or sounds of the incident between protestors and police.

- **08:26 — Switch quickly to Fox News.** We see a close-up of *Beltway Boys* host Fred Barnes (on-camera), who is riding along with the motorcade and is thus part of the diegesis even though he is close-miked and thereby constitutes a figure of authority. This is a common practice in news, wherein the broadcaster will "spatially nudge absent narrators (as sound origins) toward a spatial co-presence with events (image origins)" (Wurtzler, 1992). From a vehicle within the motorcade, Barnes says: "By and large, the crowd having a good time. Seem to be enjoying it. The motorcade is moving at a slow but steady pace, except for one moment where I think somebody said a snowball may have been thrown at the limousine or toward the president's direction. You can search people for just about everything, but nothing's going to stop them from scraping up a little snow on the side of the road and throwing a snowball. So far though, incident free, gone off just as they expected Shep." Sheppard Smith (off-camera), from the studio: "Yeah, you know … [unintelligible] I'm surprised by how few people are lining the streets there, protestors or supporters." Barnes: "I'm sorry, say it again Shep." Smith: "It's cold and miserable and all that but I'm still surprised how few people are out on the street." Barnes: "They seem to be having a great time here. The spirits are high. People are having a great time, clapping." He then talks about how people are braving the cold to get a glimpse of the ceremony. Smith shifts to

another pair of analysts (off-camera) who also discuss the crowd, offering conjecture as to why the turnout is low. What is interesting about the Fox News footage and the ABC footage is that the voices are giving completely subjective impressions of what is occurring on the street. Jennings doesn't acknowledge the protests (perhaps, in all fairness, he didn't see it) and Barnes/Smith paint a picture of harmony and celebration. The dependence on subjective analysis and description results in two very different versions of the event.

- **10:36 — Switch to MSNBC.** Here too, we hear a conversation (more of a debate) between news analysts, moderated by Chris Matthews (all off-camera). Like ABC, they are discussing the protestors. The group characterizes the protestors in the abstract as "a very vocal group here, protesting the president's policies, but they are a very vocal minority." They discuss at length the "red-blue divide" and the culture wars over religion, and separation of church and state.

- **13:18 — Switch to CNN.** Another discussion is underway about how protestors are visible "even in the non-sanctioned areas." The footage then cuts to a panning shot of two protestors. Each of the channels is clearly aware of the protest and is discussing it, but nothing is actually heard *from* the crowd.

- **14:10 — Switch to ABC.** Here we see almost the same panning shot of the two protesters as seen on CNN. Jennings: "While we watch this, let's listen to ABC's Kate Snow…" Snow gives a report on the Bush family and supporters who will meet him at the end of the procession. Jennings resumes speaking. Again the amplified voice becomes audible and Jennings cuts himself off in mid-sentence to hear. This time the voice, much nearer to the low-level ambient mics along the motorcade, is more present in the mix. Jennings: "All right, okay, so now I've finally figured out what the guy is with the microphone. The guy with the microphone is like a cheerleader, part of the inaugural ceremonies." He then stops speaking again just as we hear the amplified voice say, in very drawn out and dramatic fashion, almost like a ring announcer at a boxing match: "…George W. Bush and first lady Barbara Bush! Wave to them! Let them know that you love America!" Jennings breaks in to regain control of the situation with his signature deadpan style that serves to bolster his status as a credible news man: "Just in case

you couldn't hear it he says ladies and gentlemen let's hear it for the first lady and the president. Let's hear that you love America. And who is to say that not everybody down there doesn't love America and what it represents?" He continues to speak and the amplified sound is lowered slightly, perhaps because the motorcade mics have moved further from the voice at that point, but more likely that a live sound mixer monitoring the broadcast lowered the natural sound level so that Jennings could be heard again, thereby relegating the amplified voice back to its normal status as part of continuity sound.

- **16:38 — Switch to CNN.** There is talk about the "pre-game show, as I call it, with the Gatlin Brothers."

- **16:57 — Switch to Fox News.** Coverage here is of the eventual end point of the procession and a discussion of the celebrities and politicians who are there to greet the president upon his arrival.

- **17:17 — Switch to ABC.** ABC has assigned a reporter to discuss the earlier clash between police and protestors at the chain-link fence. Coverage is not provided through the sound of the event, but through pictures and analysis delivered by reporter in the studio. Chris Cuomo (off-camera) is providing description of the images that we can clearly see. The camera angle starts wide and then slowly zooms in closer. A fire has broken out near the fence and by this time the size of the crowd has increased. There are protesters on one side of fence and police on the other. Cuomo describes how the protesters were throwing objects that hit several police officers, and says that the protesters shouted "Shame on you." We get no sense of how Cuomo may have heard this because the ambient sound we hear is the same low-volume continuous sound coming from the motorcade several yards away and not that of the scene. This constitutes a picture-sound disconnect because the ambient sound does not match the scene. The camera cuts back and forth between the motorcade and the incident and the sound signature is continuous across the cuts. We hear very light ambient sound and no diegetic voices are heard. Even if we could hear the sounds of the protest, the mix is so low in relation to Cuomo's analysis, we wouldn't be able to get a sense of the environment.

- **18:37 — Switch to CNN.** There is no coverage—neither pictures nor sounds—of the protest. They show the orchestral band marching down the street. Commentators discuss the outcome of the recent election.

- **19:08 — Switch to ABC.** The cameras are back on the motorcade with analysts resuming their play-by-play of the procession.

The 2005 inauguration coverage clearly displays how sound serves descriptive and ideological rather than contextual purposes through analysis and description and minimal ambient coverage. The environmental sound that we do hear is extremely low-level sound that serves to provide continuity between cuts. The site of conflict in the event itself is that between President Bush and the protestors who oppose him. Rather than provide access for the public within the diegesis of the event itself, all four news channels chose to offer subjective analysis of the conflict from the "objective" distance of the studio. The camera occasionally cuts to brief pictures of individual protests, but the sound of the opposition to the ceremony is never represented. The only diegetic voice that is offered any legitimacy is that mediatized through a bullhorn. The close-miked voices, meanwhile, give background but fail to provide a voice to the wider impression of the event as it occurred on that day. This is classic television sound practice: First, each of the four broadcasters follows the exact same codes, wherein the presentation of *ceremony* dominates over public participation. Second, each delivers a program of apparent objective detachment that will meet the economic needs of a legitimate news network.

From this coverage, we can isolate five codes of sound practice used by television news broadcasters used in live televised events. They are:

1. **1920s-1940s radio aesthetic.** Disembodied, off-camera voices project an interiority of sound; clean and direct vocal tracks dominate through the use of in-studio microphones; foreground/background sound isolates the voice and projects it very prominently against other sounds.

2. **1930s-1940s film aesthetic.** Intelligibility over fidelity dominates the representation; continuity in ambient sound facilitates perceived unity through the many changes in picture perspective.

3. **Noncontextual sound approach.** Pictures show the public, including protestors, but at no time is sound heard *from* them; the public is discussed but never heard.

4. **Authoritative voices only.** All voiced content consists exclusively of nondiegetic celebrity voices—those either employed by the network or pre-selected by the network. The only non-selected diegetic voice was the one that seemed to be *mediatized* through amplification.

5. **Highly subjective analysis.** Although the code of practice was identical among the different broadcasters, the account of the same series of events was often very different. Even though protests were happening throughout, only one station (ABC) shifted from the ceremony to cover the unrest that was occurring in progress, and that was only through pictures and the words of the correspondent. In my viewing, none of the other stations reported on the incident, and Fox News went so far as to suggest that everything was "incident free."

Overall, what we find in this example and in these codes is a homogenized practice through standardization. All four broadcasters exhibited the exact same approach to sound representation. There was no deviation from the need to extract high profile, celebrity voices from the surrounding sounds, and the use of controlled ambient sound to provide an impression of reality that also serves the needs of continuity. What is particularly egregious in the close-miked voices is the content itself—a need to describe for the viewer not only what was seen but also what was *not heard*.

CHAPTER 3. THE CULTURAL AND IDEOLOGICAL IMPLICATIONS OF SOUND IN TELEVISION

The resounding ideological problem in what has been covered thus far is how sound isolates and amplifies voices of authority over those of the public, and yet does so under a guise of objective representation. Through practice and technology, television sound is enslaved to the authority of the voice, which becomes the exclusive conveyor of meaning. As such, television news sound fails in its ability to fully represent layers of context that may be available through the sound of the environment itself. In its constant need to eliminate that which may disrupt the primacy of the voice, television sound actively removes the ambient sound that might contextualize the televised representation. This is a clearly evidenced by the example of Howard Dean's speech in Iowa. As a result of this practice, television presents an artificially structured, ideological *concept* of the event as determined by the broadcaster through the voices of authority who describe it. In this section, I will argue that it is the *contextual ambient sound*—the environment and the public who participate within it—that offers a greater potential toward appreciating the nuances of culture and human behavior than authoritative voices alone can provide.

SECTION I — THE FAILURE OF OBJECTIVITY

"Almost as soon as its form is stabilized, it needs to be re-made."

- John Dewey, *The Public and its Problems*, 1927

Objectivity is an important issue to consider when exploring sound representations and context. While it remains the essential tenet of mainstream news journalism's efforts toward achieving unbiased reporting, objectivity also constitutes an ideological advantage for the broadcasting corporation. Public relations efforts aim to leverage their supposed transparency, balance, and lack of bias as a means of engendering trust on the part of an audience. Not only does this serve the economic needs of the broadcaster rather than the public good, it is also primarily delivered through spoken words. News organizations perceive their job of balance as complete when they have discussed two sides of an issue. In broadcast television, this is proffered through dialogue between two opposing ideas and delivered through two or more close-miked authoritative subjects. This promulgates the mistaken notion that understanding of a given situation can be achieved through the analysis of selectively chosen specialists narrating events using restrictive, unidirectional microphones. The outcome is that wider, more contextual representations are lost through the medium. This *ambient environment* very often includes other voices and the cultural atmosphere of a particular place that may offer a challenge to these dominant voices.

The word *objective* in its most basic, epistemological form means simply that the significance of any given situation lies in the object rather than the subject. The *Oxford English Dictionary*'s first entry for the word is: "I. Senses relating to objects, their function, and perception. 1. Of or relating to an object" (2005). The reason objective reporting is such a foundation of journalistic integrity is that the object (in this case the live event being covered) should be what emerges for the public, rather than the subject (the institution, the ideology, the reporter, or the figures of authority) unnecessarily exerting its will over the fluidity of meaning that it may or may not signify. In defining the noun *objectivity*, however, the issue becomes more problematic: "The quality or character of being objective; (in later use) *esp.* the ability to consider or represent facts, information, etc., without being influenced by personal feelings or opinions; impartiality; detachment" (*Oxford English Dictionary*, 2005). Here we find an over-reliance on rationalism that attempts to accomplish the impossible—remove the complexity of human behavior and emotion from the very human process of newsgathering, selection, and reporting.

No one better advocates this reason-based approach to objectivity than Walter Lippmann. According to Lippmann, the actions of society "can be known only by controlled reporting and by objective analysis" (1922, p. 182). Regarding news events, Lippmann suggests the following: "Unless the event is capable of being named, measured, given shape, made specific, it either fails to take on the character of news, or it is subject to the accidents and prejudices of observation" (p. 229). Lippmann's influence in the news industry was significant, and his influence is embodied in the current practice of broadcast news, which attempts to find verifiable meaning through the commentary, description and analysis of expert voices. As Lippmann says:

> And so we make our connections with the outer world through certain beloved and authoritative persons. They are the first bridge to the invisible world. And though we may gradually master for ourselves many phases of that larger environment, there always remains a vaster one that is unknown. That we still relate ourselves through authorities. (1922, p. 142)

These authority figures are the ones granted access to the medium through microphones as a means of achieving the Lippmann ideal of democratic involvement as "simple, intelligible, and easily managed" (p. 172). John Durham Peters, however, sees Lippmann as "a great critic of public discourse," (1989, p. 211) and, as a means of contrast, looks to Lippmann's contemporary, John Dewey, as an advocate for "the objectivity or publicness of *experience*" (Peters, p. 207, emphasis added). "What is needed to direct and make fruitful social inquiry is a method which proceeds on the basis of the interrelation of observable acts and their results" (Dewey, 1927, p. 36). Here we get closer to an idea of the *object* as public interaction, not as an authoritative domination over issues of meaning, facts and truth. By this definition, objective reporting is not a matter of dehumanizing the process of telling stories, but rather pledging to represent the object rather than a controlled, rational, subjective impression of it. It is an understanding that the meaning of an event is more often than not arbitrary and its significance is best determined by the public (the object) rather than the broadcaster (the subject). In the process, this gives the event over to the public who view and hear it to decide for themselves what its meaning may be.

As witnessed in the practice of journalism, Lippmann's notion of objectivity is not only flawed, but dangerous. It invites opportunities to manipulate the meaning of events by those in power. If, by contrast, we take Dewey's approach to the world as not one of facts but one of experience, we find a wider model of representation for news sound practice to follow. This mode of presenting a more direct experience of the sound of the environment-as-object has a greater opportunity to transcend containable facts and better reflect the interaction of society through the sound of those interactions within the place in which they occur. Through the capturing of surrounding voices and the cultural significance of *place*, dominant voices of authority are contextualized and a broader sense of cultural awareness becomes more possible.

The question to ask, therefore, is a common one in the history of thought on mediated communication: "Can the object itself ever break through the veil?" (Peters, 1999, p. 204). We make the mistake of believing that objectivity is only available through language, when objectivity can instead refer to any means toward contextualization. And context is nothing more than representing an event in its larger and more open form.

Language carries with it inherent biases and ambiguity; it is exclusive more often than it is inclusive, particularly when the texts are written in accordance to the commercial aspirations of the news-making agency. Efforts toward representing the object more than the subject are efforts toward contextualizing the language of individuals with the communal environment in which ideas, interactions, and events occur. It is a means of providing the public contextualism through sound by offering a multiplicity of voices interacting in the environment in which they occur. Not only is the wider public represented, but the television public can then form opinions and take action based on the event more than the messenger. Ambient sound—the sound of community, of culture, and of place—can represent this multiplicity of ideas and a wider scope of understanding, and in so doing, television news can achieve more open representations of events.

SECTION II — REPRESENTATIONAL AND REPRESENTATIVE SOUND

"The dust is part of the case: the caustic dust carried by the vile wind. Democracy, truth, art, equality, culture: all these we carry in our heads, but, in the street, the wind is everywhere."

- Raymond Williams, *Culture & Society: 1780-1950*. (p. 285)

Democracy, as John Dewey suggests, "is a word of many meanings" (1927, p. 82). There is the strictly political definition as relating to a public form of government in which citizens in society contribute to its structure. There is also the more general concept of democracy as social inclusion. By this second definition we can take the model of democratic representation and apply it to the cultural industry of televised representation. With television, particularly in regard to television sound, we find a nonrepresentational entity at work. When Dewey writes about the problems of public representation in the *state,* we can also find relevance to the way television sound is dominated by structures of control and authority over the public it purports to serve:

> They represent an effort in the first place to counteract the forces that have so largely determined the possession of rule by accidental and irrelevant factors, and in the second place an effort to counteract the tendency to employ political power to serve private instead of public ends. (1927, p. 83)

This is the struggle in which television sound operates. First, from the institutional/ideological perspective, it is a structure that is exclusive and closed, one that is dictated by power elites. The only people with access to a message are those who pass the Lippmann standard, the so-called "credible sources," the ones the authority of the industry itself has selected to represent the views of society. These are the only voices granted access to the public discourse through television news broadcasts. Second, and as a result, television sound is not representative of the public it is entrusted to serve. All efforts toward sound are aimed at enhancing the message of authority figures—and by extension, serving the commercial interests of the broadcaster—over the needs of the public. The ideological undercurrent of this is that it wraps an illusion of pluralism through a notion of objectivity that the public will accept.

63

While one could argue successfully that it is the voices of television that provide the necessary discourse that is needed in a democracy, what happens in practice is that *selected* voices uphold the authority of the institution and exclude others, uncoached and unselected, who may have something to add to the discussion. The discourse we find on television is not one of voices of the people. With its restrictive practice of sound gathering, television dialogue is too stringently planned and controlled to provide access to those participating in the action as it happens. Television therefore fails to engage the public because conversation only occurs among those the institution deems worthy of speaking. Ultimately, television fails by confusing chatter with discourse. Instead of *talking over* the event, television should use sound to *offer the experience* of the event.

While much of the scholarly work on media centers on the *representational* aspects of sound, the discussion should also open up to include more consideration of the *representative* issues of sound. The matter is one of examining not only the media system, but also the public it promises to reflect. The question to ask is: How can sound better include those who are not granted access to one another through the medium? I am not suggesting that we offer microphones to every person present in an event. This would simply be a continuation of restrictive practice, not to mention an increase in noise. I'm suggesting instead that the bandwidth of sound be widened to take on more points of reference. This widening effort can be aided by more inclusive microphone types that are sensitive to environmental context, and also by taking advantage of new forms of television broadcasting. For example, digital television broadcasting (DTV) offers a significantly more open playback system with digital surround sound. Surround sound enables the capturing of multiple sources of audio (more coverage of an environment) into discrete playback channels (multiple home speakers) in such a way that does not adversely affect the intelligibility of close-miked dialogue sources. As Tomlinson Holman notes in regard to surround sound playback: "With more channels operating, there is greater likelihood that multiple streams can be followed simultaneously" (Holman, 2000, p. 117).

The problem in sound representation is not that language is the wrong form of sonic content in television news, but that language is problematic when it is manifested as

an agenda dictated by controlled forces and delivered in a way that is devoid of audible context.

> Dialogue can be a wonderful method for enforcing imagination of the other's position and is obviously a far superior mode of handling differences than fisticuffs or nerve gas, but it is not in itself an adequate communicative vehicle for bearing the full varieties of moral experience. (Peters, 1999, p. 160)

A television soundtrack *can*, however, contextualize both pictures and sound by presenting the sound of experience, and in the process, provide an audience with broader access to an event. Ambient sound representations achieve a middle ground between language and pictures. Neil Postman and Steve Powers point to the two elements at play in television news—pictures and words. "[I]ndividual pictures give us the world as object; language, the world as idea" (1992, p. 105). Ambient sound exhibits elements of both. It reflects the object at the same time that it suggests the possibility of ideas that have traditionally remained hidden beneath the surface.

Ambient sound is also a means by which independent media can distinguish itself from the mainstream. Not tied down to institutionalized rigors of "proper" practice and expensive narrow-focus microphones, independent media has the ability to offer broader representations of sound as a more inclusive alternative, just as Joe Jensen's footage of the Dean rally did. This is especially true in the current environment of cable and online media. News footage no longer has to be dictated by a select few network news corporations. If we could see a change in Federal Communications Commission protocols that is more welcoming to independent or public forms of broadcasting, the opportunity will arise for moving beyond these restrictive modes of practice. Even without the cooperation of cable, the Internet has the ability—as evidenced by Joe Jensen's donation of his material to valuejudgment.org—to be the alternative conduit of information. Hopefully, we will see more efforts toward contextualizing the dominant news text itself, wherein the major news media show their restrictive form of coverage while the Internet shows the same event filmed by a camcorder with a more open, omnidirectional microphone and a better consideration of contextual sound. In the process, this alternate version might force more responsible sound practices within the mainstream. Perhaps through more televised moments that capture multiple perspectives on events, an

openness in coverage will fold back into the mainstream practices of standardized news gathering and reporting.

CONCLUSION

Television news sound is a hierarchical practice in which the voices of experts are preferenced over the sensual experience of events because such practice offers an appearance of legitimacy and objectivity. Unfortunately, sound in television news and live event coverage—the very coverage that should be providing a sense of openness—is exclusionary and restrictive in nearly every sense. From sound's early history, it has served as an agent inferior to the image, one whose role was to generate capitalist interest. We can see how today's sound carries ideological functions in bolstering the illusion of continuity and upholding the voices of authority in news reporting, serving a commercial need for informational clarity over inclusion of sonic context, pushing the technical practice of manageable predictability over acoustic environments, and promoting the technological development of systems and devices that supported each of these aims. The history of sound—which defines, in essence the current practice of sound in television—is a story of people, the technicians and artists themselves who have accepted the directives of those agents which have little interest in sound itself. Those outside the field of sound do not respect it as an art or even as a technique. Instead, sound has always been the tool of some other manipulative purpose. Sound compensates for problems the picture cannot repair and as a result, injects legitimacy into mediated messages. It gives authority to a handpicked selection of agents who wield it as an ideological tool. The outcome of this practice is that dominant perspectives always prevail in broadcast news. These voices uphold the legitimacy of the broadcaster by meeting its commercial interests. Broadcast news, as a result, does not engender an interactive, open, multicultural flow of multiple ideas. Rather it constitutes a flow of *capital*, wherein a particular *impression* of information and experience is packaged as a product for consumption. So when seeking an answer to why Howard Dean sounded the way he did over the medium of broadcast news, and why that particular sound was replayed with such fascination, we find an answer in the commercial inclinations of the broadcasting entities. The primacy of the voice is *built into* the system of broadcasting— it is encoded into the technology (the microphone) and the institution (the news media).

Such encoding preserves a manageable product for consumption. The unfortunate outcome of such practice is a narrowing of ideas.

This is not to say that content does not require organization. To some indefinable degree, it is in the structuring of sound that communication happens. So it is not simply a matter of opening up the medium to every sound that may pass through it, rather that we should take advantage of what sound offers as a means of providing access to the wider scope of ideas within an environment. This can be done by opening up the sonic landscape to include the ambient sound of experience and community, while *also* maintaining clarity in voices that speak about it. It is a matter of contextualizing the cold, inhuman distance of authoritative analysis and description with the public sphere of human activity. We need to be able to hear our world before we can talk about what it means.

This contextualization is available through the capturing of wider sonic environments. Closely miked subjects through unidirectional microphones have their place in reproducing intelligibility in dialogue. But the remainder that is commonly left out can be brought back through additional channels devoted to capturing environmental representations. Videotape has the ability to capture two channels of sound, and therefore can include a directional microphone (lavaliere or boom) for voices and an omnidirectional design that is devoted to capturing as much of the surrounding environment as possible. "An omni placed in any size room will give a sense of the actual space at normal levels because it mixes a balance of background and foreground ambience with the principal signal (within its maximum area of sensitivity)" (Zaza, p. 226). With the use of a common three-channel mixer, yet another channel is available for representing ambient sound. Sound practitioners and camera operators also need to be sensitive to the larger context of every event being covered. Particularly in the age of emerging digital broadcasting, with increased fidelity and a wider bandwidth available through surround sound broadcasting, the opportunities are there for representing a more inclusive sonic representation. While this may require extra time in the field, as we have seen, the ethical considerations are worth the effort.

Television has a unique potential for public good. No other form of media can convey television's combination of audiovisual representation with the immediacy of

events unfolding in real time. But television fails in its potential to open up to the richness of experience that comes from the sound of culture, or of the interaction of human activity in its environment. By including the ambient contextual sound in the broadcasting of our shared history as it happens, two critical goals are accomplished: (a) the television viewing public is granted the opportunity to hear for themselves the sound of an event as it occurs rather than relying on subjective descriptions by voices of authority, and (b) the public being *represented by* television in its coverage of events is heard more openly. In both situations, the public is given access to that which has been denied by restrictive practice. It is the audible equivalent of widening the camera frame without losing focus. Ambient sound, if practiced with good intentions, gives the audience an answer to the question: "Yes, I hear what you're *describing*. But would you mind also letting me hear it for myself?"

REFERENCES

Alten, S. (1990). *Audio in media*. (3rd Ed.). Belmont, CA: Wadsworth.

Altman, R. (1985). The technology of the voice (part 1). In *Iris (3)*1. (pp. 3-20).

Altman, R. (1986). Television sound. In Modleski, T. (Ed.), *Studies in Entertainment*. Bloomington, IA: Indiana University.

Altman, R. (1992a). The material heterogeneity of recorded sound. In R. Altman (Ed.), *Sound Theory Sound Practice* (pp. 15-34). New York: Routledge.

Altman, R. (1992b). Sound space. In R. Altman (Ed.), *Sound Theory Sound Practice* (pp. 46-64). New York: Routledge.

Altman, R. (1994). Deep-focus sound: Citizen Kane and the radio aesthetic. *Quarterly Journal of Film & Video 15*(3). (pp. 1-33).

Altman, R. (2005). *Silent film sound*. New York: Columbia University.

Associated Press. (2004). "CNN says It overplayed Dean's Iowa scream". [Retrieved from ABC Action News Tampa, Fl website at http://www.abcactionnews.com/entertainment/stories/0402/040209cnn.shtml on July 14, 2005.]

Auslander, P. (1999). *Liveness: Performance in a mediatized culture*. London: Routledge.

Borwick, J. (1990). *Microphones: Technology and technique*. London: Focal Press.

Bracken, C. C. & Atkin, D. J. (2004). How screen size affects perception of television: A survey of presence-evoking technology in our living rooms. *News Photographer 59*(4). (pp. 23-28).

British Broadcasting Corporation. (1951). *Microphones*. London: Iliffe & Sons.

Bruck, J., Grundy, A. and Joel, I. (1999). *An audio timeline: A selection of significant events, inventions, products and their purveyors, from cylinder to DVD*. (Unpublished Audio Engineering Society compendium). (Retrieved on July 4, 2005 from http://www. aes.org/aeshc/docs/audio.history.timeline.html).

Campbell, R. (1987). Securing the middle ground: Reporter formulas in *60 Minutes*. *Critical Studies in Mass Communication 4*. (pp. 325-350).

Carpignano, P., Andersen, R., Aronowitz, S. & DiFazio, W. (1990). Chatter in the age of electronic reproduction: Talk television and the 'public mind'. *Social Text 25*(26). (pp. 95-120).

Center for Media and Public Affairs, the. (2004). "Network news focus: Flubs, fluff—not functional". (Feb. 9, 2004). Retrieved on July 22, 2005 at http://www.cmpa.com/pressReleases/NetworkNewsFocus.htm

Chion, M. (1994). *Audio-vision: Sound on screen*. New York: Columbia University.

Chion, M. (1999). *The voice in cinema*. Trans. by Gorbman, C. New York: Columbia University.

Compesi, R. J. (2000). *Video field production and editing*. (5th Ed.). Boston: Allyn and Bacon.

CNN. (2004, January 20). [Television broadcast video archive]. Retrieved November 16, 2004 at http://www.cnn.com/2004/ALLPOLITICS/01/20/elec04.prez.dean/index.html

C-SPAN. (2004, January 19). Howard Dean reaction to Iowa caucus results: Former Vermont Governor Howard Dean reacts to the results of the Iowa caucuses. [Television broadcast video archive]. Retrieved November 4, 2004 at rtsp://video.c-span.org/project/c04/c04011904_dean.rm

Dayan, D. & Katz, E. (1992). *Media events: The live broadcasting of history*. Cambridge, MA: Harvard University.

Dewey, J. (1927). *The public and its problems*. Athens, OH: Ohio University.

Doane, M. A. (1985a). Ideology and the practice of sound editing and mixing. In E. Weis & J. Belton (Eds.), *Film sound: Theory and practice* (pp. 154-161). New York: Columbia University.

Doane, M. A. (1985b). The voice in the cinema: The articulation of body and space. In E. Weis & J. Belton (Eds.), *Film sound: Theory and practice* (pp. 163-176). New York: Columbia University.

Eargle, J. (2001). *The microphone book.* Boston: Focal Press.

Fine, D. (2004). "On second thought: Maybe there's a reason the media distorted the 'I Have a Scream' speech". *Fort Worth Weekly Online*. (March 3, 2004). Retrieved June 29, 2005 at http://www.fwweekly.com/issues/2004-03-03/thought.html

Fischer, L. (1985). Applause: The visual and acoustic landscape. In Weis, E. and Belton, J. (Eds.) *Film Sound: Theory and Practice*. (pp. 232-246). New York: Columbia University.

Gans, H. J. (1979). Deciding what's news: Story suitability. *Society 16*(3). (pp. 65-77).

Graham, T. (2004). Media-powered Howard: ABC jerks their knees to apologize and sympathize with the Dean scream. *National Review Online*. 1/30/2004. Retrieved on November 21, 2004 at http://www.nationalreview.com/comment/tgraham200401300919.asp

Handzo, S. (1985). Appendix: A narrative glossary of film sound technology. In Weis, E. and Belton, J. (Eds.) *Film Sound: Theory and Practice*. (pp. 383-426). New York: Columbia University.

Holman, T. (2000). *5.1 surround sound: Up and running*. Burlington, MA: Focal Press.

Kellogg, E. W. (1955). History of sound motion pictures: Second installment. *Journal of the SMPTE 64*(7). (pp. 356-374).

Lastra, J. (2000). *Sound technology and the American cinema: Perception, representation, modernity*. New York: Columbia University.

Lippmann, W. (1922). *Public opinion*. New York: Free Press.

Louw, E. (2001). *The media and cultural production*. London: Sage.

Metz, C. (1985). Aural objects. In E. Weis & J. Belton (Eds.), *Film sound: Theory and practice* (pp. 154-161). New York: Columbia University.

Murch, W. (2001). *In the blink of an eye: A perspective on film editing*. (2nd Ed.). Los Angeles: Silman-James.

Oxford English Dictionary. (2005). Oxford, UK: Oxford University Press. Retrieved online at: http://0-dictionary.oed.com.opac.sfsu.edu/

Peters, J. D. (1989). Democracy and American mass communication theory: Dewey, Lippmann, Lazarsfeld. *Communication 11*(3). (pp. 199-220).

Peters, J. D. (1999). *Speaking into the air: A history of the idea of communication*. Chicago, Il: The University of Chicago.

Pitkin, W. B. and Marston, W. M. (1930). *The art of sound pictures*. New York: D. Appleton & Co.

Postman, N. and Powers, S. (1992). *How to watch TV news*. New York: Penguin.

Sawyer, D. (2004). The Dean scream: Diane Sawyer looks at the version of reality you didn't see on TV. ABCnews.com. 1/29/2004. Retrieved on Nov. 21, 2004 at http://abclocal.go.com/kgo/news/012904_nw_dean_scream_abcnews.html

Shaw, B. (Reporter). (1991, January 16). [Live news broadcast]. Atlanta: Cable News Network.

Shrivastava, V. (1996). *Aesthetics of sound: Critical analysis of sound design in television and motion pictures*. Dubuque, IA: Kendall/Hunt.

Shone, l. w. (2004, February 10). [Msg 15]. Message posted to http://www.valuejudgment.org/archives/000441.html

Sonnenschein, D. (2001). *Sound design: The expressive power of music, voice, and sound effects in cinema*. Studio City, CA: Michael Wiese Productions.

Sterne, J. (2003). *The audible past: Cultural origins of sound reproduction*. Durham, NC: Duke University.

Jensen, J. (Producer). (2004). The real 'Dean scream' video. [Video footage]. Retrieved Nov. 17, 2004 at http://www.valuejudgment.org/archives/000441.html

Wayne, M. (1997). *Theorizing Video Practice*. London: Lawrence & Wishart Limited.

Weis, E. (1995). Sync tanks: The art and technique of postproduction sound. *Cineaste 21*(1). (electronic version). Retrieved on Oct. 11, 2004 at http://www.filmsound.org/synctanks/

Welles, O. (Producer). (1938). *The war of the worlds* [Radio broadcast]. United States: Columbia Broadcasting Corporation. Script retrieved at http://members.aol.com/jeff1070/script.html

Wente, E. C. (1917). A condenser transmitter as a uniformly sensitive instrument for the absolute measurement of sound intensity. *Physical Review 10*(1). (pp. 39-63).

Williams, R. (1958). *Culture and society: 1780-1950*. New York: Columbia University.

Wurtzler, S. (1992). She sang live but the microphone was turned off: The live, the recorded and the subject of representation. In R. Altman (Ed.), *Sound Theory Sound Practice* (pp. 87-103). New York: Routledge.

Zaza, T. (1991). Improving TV audio. (Ch. 9). In *Audio Design: Sound Recording Techniques for Film and Video*. (pp. 220-244). Englewood Cliffs, NJ: Prentice Hall.

Zettl, H. (1999). *Sight sound motion*. Belmont, CA: Wadsworth.

FOOTNOTES

1 I use the word "events" in a broad sense to encompass any occurrence *covered by* news media, rather than the more narrow sense of pre-planned "media events" that are designed specifically for television programming.

2 See Sterne, J. (2003) for the former, and Lastra (2000) and Altman (1992b) for the latter.

3 See Altman, R. (2004). *Silent Film Sound.* New York: Columbia. for an in-depth look at the "silent film era," a period of cinema that was by no means devoid of sound.

4 More recently, the FCC has proposed the adoption of the Digital Television (DTV) protocol, which allows for the broadcasting of high definition digital audio and 5.1 digital surround sound, but it has not received wide enough use to warrant analysis at this time.

5 Which she did. See Graham, T. (2004). *Media Powered Howard.*

6 This diegetic/nondiegetic relationship will come into play in the upcoming chapter, which includes the televised coverage of the presidential inauguration.

7 Here, "live televised event" is more aligned with Dyan & Katz's concept of the "media event," and as such is distinct from the previous use of "event" as an occurrence that is covered by a news institution.