

The

Sound Studies Reader

Edited by

Jonathan Sterne

First published 2012
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2012 editorial and selection matter Jonathan Sterne; individual chapters, the contributors

The right of Jonathan Sterne to be identified as editor of this work has been asserted by him in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data

The sound studies reader / edited by Jonathan Sterne.

p. cm.

Includes bibliographical references and index.

1. Sound—Recording and reproducing—History. 2. Sound—Recording and reproducing—Social aspects. 3. Hearing. 4. Listening. I. Sterne, Jonathan, 1970-

TK7881.4.S684 2012

621.389—dc23

2011052981

ISBN: 978-0-415-77130-6 (hbk)

ISBN: 978-0-415-77131-3 (pbk)

Typeset in Baskerville
by RefineCatch Limited, Bungay, Suffolk

James Lastra

FIDELITY VERSUS INTELLIGIBILITY

THE TWO GENERAL MODELS of sound recording that dominated—and to a remarkable degree continue to dominate—technicians’ ideas about sound representation could be called the “phonographic” (or “perceptual fidelity”) model and the “telephonic” (or “intelligibility”) models. Like its nineteenth-century predecessor, perceptual simulation, the 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. A recording of an orchestra, for example, should try to preserve the reverberant space within which the sounds were produced, while a telephone, being designed with a very different social function in mind, would never be suitable for this. However, a “phonographic” recording could just as easily represent one of the “worst” seats, while still remaining “faithful” or specific. A telephone, in contrast, sacrifices acoustic specificity in favor of rendering speech clearly under widely varying conditions.

For each model articulated by sound engineers and researchers, the very conception of what a sound “is” differs in important ways. The “fidelity” approach assumes that all aspects of the sound event are inherently significant, including long or short reverberation times, ratios of direct to reflected sound, or even certain peculiarities of performance or space. The “telephonic” approach, not literally limited to telephones and voices, assumes that sound possesses an intrinsic hierarchy that renders some aspects essential and others not. As with our academic theorists, the relevant terms of comparison are uniqueness versus recognizability or event versus structure, and consequently, these terms presume different ideals of “good” representation.

Within the classical Hollywood style, we can find instances of both the “telephonic” and “phonographic” approaches to sound 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 the dialogue.¹ Even when a speaker appears to turn away, a high level of direct sound often implies that he or she is still speaking “to” the auditor, because speech is understood not simply as an abstract sound but as a sound with a specific social and narrative function.² This statistically dominant form of recording differs markedly from the less common technique that has been dubbed point-of-audition (POA) sound, which tends to correspond more closely to the phonographic model of represented hearing.

In spite of its relative rarity as a technique, the phonographic or simulation model ruled the theoretical roost. This should not surprise us, since it lent itself easily to the reigning theoretical orthodoxies of the day. By insisting on a physically real observer as the principle of representational coherence, it was easily understood as a variation on the “invisible witness” model of narration. From Hugo Münsterberg and V. Pudovkin to André Bazin, classical film theorists have tried to account for narration by claiming that editing (for example) mimicked the perceptions of an “invisible observer.” In this view, edits reflect the shifting attention of a real witness, and filmmaking becomes the process of moving that witness around through space to the most salient details of a scene. The frequently impossible spatial leaps and changes of perspective characteristic of standard continuity editing give the lie to this model, but it persisted nonetheless. As David Bordwell notes, this theoretical expedient allowed filmmakers and critics to collapse the observer with both the camera and the narrator. Furthermore, the ambiguity of the term “point of view” (POV) allowed the conflation of both its optical and intellectual senses. Thus, he argues, “the invisible-witness model became classical film theory’s all-purpose answer to problems involving space, authorship, point of view, and narration.”³ Whatever inconsistencies (sometimes a real observer, sometimes an “ideal” one, sometimes an “omniscient” one, etc.) or outright silliness the invisible observer produced, the model died hard.

In addition to its conceptual and rhetorical benefits, the invisible observer’s insistence on the foundational importance of POV responded to other pressures. If we look back to the nineteenth and early twentieth centuries and remember the extent to which the technical media disrupted customary understandings of the relationship between perception and representation, the model makes even more sense. While a filmmaker like Dziga Vertov might celebrate the machine-vision produced by the camera and editing, and extol its modernity and superhuman capacities, the experience of radical shifts of space and time made possible by the cinema (and even the gentler dislocations of continuity editing) presented the spectator with an unsettling and profoundly inorganic experience of the visible world.

Even a conventionally edited scene might offer microscopic or bird’s-eye perspectives, and it would almost certainly jump back and forth across the shoulders of couples engaged in intimate conversation. Even if not literally shocking, this was a novel and potentially confusing experience. By assimilating these obviously inhuman processes to a thoroughly (if problematical) human model, critics, theorists, and classical filmmakers carved a space for subjectivity in film narration and “anthropomorphized” its more perplexing possibilities. In essence, POV “humanized” machine perception. In a larger sense, by focusing on character, adjusting framing to the human body, and emphasizing psychological interiority and character-motivated

POV, the classical cinema systematically worked to minimize the more disturbing tendencies of the new medium. It is therefore even less surprising that variations on this “anthropomorphism” would creep into technical discussions.

POA sound (like the POV shot) attempts to represent the experience of hearing within the diegesis, normally the hearing of a character. Sometimes this is indicated by muffling sounds, as if hearing them through a wall, by including a Doppler effect to indicate the rapid passage of a sound source (by an increase or decrease in volume indicating the approach or retreat of the source), or perhaps by giving a clear sense of the acoustics specific to a particular location. All of these different characteristics can be assimilated under Altman’s term “spatial signature,” which includes all those indicators of the spatial and temporal specificity of sound production and reception that characterize any recording as unique, and that create an effect of simulated perceptual presence.⁴ Both the POV shot and POA sound represent spaces that are to be taken as diegetically “real,” and as heard by an embodied perceiver within that space. It therefore implicitly confirms the classical cinema’s emphasis on the human character, while simultaneously rendering its other historical norms more apparently “human.” In a rather straightforward fashion, it simultaneously embodies the idea of representation as perceptual simulation and responds to the same pressures that elicited its appearance in the nineteenth century. The difference between a sonic space whose principal goal is the intelligibility of some sounds at the expense of others (foregrounding narratively important information against a reduced, generic background), and a space that is constructed in order to represent a particular real act of audition, embodies the basic difference between the telephonic and phonographic models.

In a typical Hollywood film, visual space, when not explicitly marked as a POV shot, is primarily a marker of the narrative or enunciative level of the presentation. In other words, its dimension, angle, framing, etc. are typically not understood as someone’s actual vision (neither narrator nor character), but as an index of narrative emphasis. Unlike the POV shot, which is marked as a unique perception in a specific, diegetically real space, a typical shot can usually be replaced by another shot that represents a noticeably different view of the space. A typical shot’s index of substitutability is much higher than that of a POV shot since it presents *narratively important* information rather than *perceptually specific* information. It is indicative, however, that sound technicians who call for perceptual fidelity in sound and image representation often call for an “invisible witness” model of cinematic narration as well. Such a heuristic seems obviously to encourage one to think of filmmaking as a stringing together of unique perceptions.⁵ The sonic specificity implied by the standard of perceptual fidelity was quickly pressed into a carefully defined, and carefully delimited, role. It was soon harnessed almost exclusively to particular characters within a fictional world.

Sound recorded as if heard by a character—POA sound—became the primary instance within the classical system where the spatial characteristics of sound might manifest themselves (often, significantly, for narrative purposes). This however, was the exceptional case. In general, close frontal miking of actors, which minimizes reflected and indirect sound, became the norm for dialogue since, as one researcher put it, “In no case did an increase in reverberation cause an increase in articulation

[i.e., intelligibility].”⁶ Or as another phrased a related observation, “The quality of a recording is effectively independent of the reverberant characteristics of the set if the microphone is within approximately three feet of the speaker.”⁷

Together, the two quotations indicate that in certain contexts a “good” recording could be associated with recognizable speech sounds, and dissociated from any strict sense of fidelity.⁸ In fact, some researchers from the period began to argue that perceived “naturalness” seemed to be correlated with the presence of reflected sound, while “articulation” was associated with direct sound.⁹ In other words, these articles suggest that indications of spatial specificity and naturalness were linked to certain forms of *unintelligibility*, or at least suggest the possibility that fidelity and intelligibility are not necessarily related. Although my argument here focuses heavily on speech, I would suggest that terms such as “intelligibility” have analogs in the case of sound effects as well, where something like “recognizability” or “identifiability” performs roughly the same function—that of comprehension.¹⁰ A recording with a high degree of reflected sound, or some other indicator of spatial signature and temporal specificity, corresponds to an approach that considers sound an *event*, while closely miked sound, with a relatively “context-less” signature, corresponds to sound considered as an intelligible *structure*—as a signifying element within a larger system.

According to a 1931 article by Carl Dreher, there need not be any irresolvable conflict between the two approaches, however.

Since the reproduction of sound is an artificial process, it is necessary to use artificial devices in order to obtain the most desirable effects. For example, it is normal procedure to reproduce dialog at a level higher than the original performance. This may entail a compromise between intelligibility and strict fidelity.¹¹

Clearly at stake are not the sounds “themselves,” but as Alan Williams says, the signs of those sounds—even in the case of supposedly strictly accurate recordings. Nevertheless, the problem of strict fidelity arises again and again in other contexts. John L. Cass, for example, complains that the “illusion” is destroyed when a spectator becomes aware that despite changes in shot scale, the recordist has maintained “close-up” sound in order to ensure intelligibility, thereby violating the presumed norms of “sound perspective.” Since “the resultant blend of sound may not be said to represent any given point of audition,” the spectator is left feeling that he is experiencing the “sound which would be heard by a man with five or six very long ears, said ears extending in various directions.”¹²

Here Cass is, of course, assuming that film narration works by the “invisible observer” method, stringing together a series of unique and spatiotemporally real perceptions. However, filmic narration, as Bordwell points out, does *not* need the supposition of a flesh-and-blood observer to perceive every element *for us*.¹³ “Real” space, as it is perceived in actuality, is simply not an issue for most images on the screen: their scale and angle are functions of narrative emphasis, not of more or less precise perception. Cass’s symptomatic insistence on the model of narration as perception, however, indicates the extent to which a model of precise perceptual duplication ruled the imaginative world of film professionals, helping to shape the

practices of sound representation for decades to come, even if it was ultimately to be marginalized or rejected. A legacy of earlier encounters with technology, the illogical but persistent “invisible auditor” served to rehumanize the cinema when it most threatened to become inhuman. The resilience of this persuasive but misleading model also illustrates how reigning notions of filmic representation may come into conflict with institutional and industrial demands.

Notes

1. It might be argued that, ultimately, “voices” (be they Al Jolson’s, Bert Lahr’s, or Mischa Elman’s violin) were the only sounds that Hollywood had an inherent *need* to reproduce accurately in synchronization. Although the theoretical discourse promoted the importance of recording music (an emphasis derived from the Vitaphone’s “disc” format?), almost every “problem” with the apparatuses involved either synchrony or dialogue—rarely if ever inaccurate rendering of instrumental timbre.
2. Although in the Vitaphone shorts (e.g., the *Tannhäuser* overture) the moving violin bow, the cymbal crash, and various other visible manifestations of sync replace the moving lips, it is not difficult to see the analogy between these types of sounds and speech.
3. Recorded musical accompaniment for films such as *Don Juan* (1926) and *Old San Francisco* (1927) was important to the extent that it further standardized the commodity, but the added cost to producers could not have been an effective inducement to convert to sound. The musical and sound effects accompaniment typical of both films was not necessarily any better synchronized than that provided by a gifted theater orchestra, so novelty was minimal. Warner stood to profit most through the presentation of vaudeville and concert acts who were salable precisely by providing their signature sounds in a manner which left no doubt that the persons depicted on the screen were producing those sounds. Concern with sound source was prominent in Bell Labs’ Public Address research as well. See, e.g., T. A. Dowey, “Public-Address Systems,” *Bell Laboratories Record* 3.2 (October 1926): 50–56.
4. Aside, that is, from certain forms of (narrative-driven) ethnic accent. The importance of these deviances from the King’s (or rather Bell Labs’) English is their ability to provide motivations for certain forms of melodramatic narrative. Warner’s *Old San Francisco*, *Noah’s Ark* (1929), and *The Jazz Singer* (1927) all attest to this use of accent (the first only in intertitles). Moreover, such accents rarely if ever impede dialogue intelligibility, while they often provide important character information and plot complication.
5. See Rick Altman’s discussion of “for-me-ness” in “The Technology of the Voice” (Part 1), *Iris* 3.1 (1985): 3–20.
6. David Bordwell, *Narration in the Fiction Film* (Madison: University of Wisconsin Press, 1985), 9.
7. Rick Altman, “Sound Space,” in Altman, ed., *Sound Theory/Sound Practice* (New York: Routledge), 46–64.
8. Interestingly, the tendency to think of sound representation as records of perception duplicates the early tendency to treat the individual shots of, say, a chase film as autonomous wholes whose unity is provided by point of view rather than a higher-level narrative unity. See Tom Gunning, “Non-Continuity, Continuity, Discontinuity: A Theory of Genres in Early Film,” *Iris* 2.1 (1984): 101–12; Charles Musser, “The Travel Genre in 1903–4: Moving Towards Fictional Narrative,” *Iris* 2.1 (1984): 47–60; and André Gaudreault, “The Infringement of Copyright Laws and Its Effects (1900–906),” *Framework* 29 (1985): 2–14.
9. J. C. Steinberg, “Effects of Distortion Upon the Recognition of Speech Sounds,” *Journal of the Acoustical Society of America* (hereafter, *JASA*) 2.4 (October 1930): 132.
10. Franklin L. Hunt, “Sound Pictures: Fundamental Principles and Some Factors Which Affect Their Quality,” *JASA* 2.4 (April 1931): 482.

11. For example, Hunt, "Sound Pictures," 481–82. It also makes clear that very "normal" kinds of sound recording paid no heed to the credo which claimed that the "eyes and ears" of the recording apparatuses had to be linked like those in a real body. It further indicates that at least a portion of the technicians involved in research around problems in sound film were aware that Hollywood was primarily in the business of selling narratives, not optical and acoustic effigies of real experience. For a good example of a technician grappling with the pair intelligibility/naturalness, see Carl Dreher, "Recording, Re-recording, and Editing of Sound," *Journal of the Society of Motion Picture Engineers* 16.6 (June 1931): 756–65.
12. Hunt, "Sound Pictures," 499.
13. A notable exception are those sound effects that require a strong dose of spatial acoustics in order to be genetically recognizable. Hence we often get sound/image scale mismatches which do not disturb us in the least because the rendering of the sound stresses its "name" rather than its actuality.