



# MIT Open Access Articles

*Acoustics of the one person space: headphone listening, detachable ambience, and the binaural prehistory of VR*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

<b>Citation</b>	Roquet, Paul. "Acoustics of the one person space: headphone listening, detachable ambience, and the binaural prehistory of VR." <i>Sound Studies</i> 7, 1 (April 2020): 42-63 © 2020 Informa UK Limited
<b>As Published</b>	<a href="http://dx.doi.org/10.1080/20551940.2020.1750270">http://dx.doi.org/10.1080/20551940.2020.1750270</a>
<b>Publisher</b>	Informa UK Limited
<b>Version</b>	Author's final manuscript
<b>Citable link</b>	<a href="https://hdl.handle.net/1721.1/130242">https://hdl.handle.net/1721.1/130242</a>
<b>Terms of Use</b>	Creative Commons Attribution-Noncommercial-Share Alike
<b>Detailed Terms</b>	<a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a>

# Acoustics of the One Person Space: Headphone Listening, Detachable Ambience, and the Binaural Prehistory of VR

Paul Roquet

Citation: Paul Roquet (2021) “Acoustics of the One Person Space: Headphone Listening, Detachable Ambience, and the Binaural Prehistory of VR,” *Sound Studies*, 7 (1): 42-63.

Published version available online at <https://doi.org/10.1080/20551940.2020.1750270>

This article traces forgotten debates over the cultural acceptability of headphone listening in order to present an audio-centric history of private virtual space. I focus on how cultural attitudes towards headphone use transformed in tandem with developments in 3D binaural audio, new practices of solitary listening, and sound engineers’ attempts to solve the “sound in the head” problem accompanying headphone use. Extending Jonathan Sterne’s work on the “detachable echo” of modern sound mixing to what I call the *detachable ambience* of head-mounted spatial audio, I explore what happens when headphone listening eliminates the mediation of the surrounding physical space to create a solitary virtual acoustics. Examining the post-Walkman normalization of headphone use, I argue it was only with the architectural emergence of what Nango Yoshikazu calls the “one person space” that headphone listening became acceptable as a personal media practice, setting the stage for more recent debates over the privatized three-dimensional space of the virtual reality head-mounted display.

Keywords: headphones, binaural, acoustic space, virtual reality (VR), detachable ambience, Japan

While the peep box and the stereoscope are often mentioned as predecessors of the virtual reality (VR) head-mounted display, audio technologies offer a more direct precursor. As Steven Jones (1993) and Frances Dyson (1996) noted during the first VR boom, even though the discourse surrounding virtual reality often emphasizes the visual register, much of the 360-degree immersion promised by VR was first achieved in sound. These sonic precedents for virtual immersion were not lost on some of the central players in VR history. William Gibson's notion of "cyberspace" was inspired in part by his experiences with the Sony Walkman (Headlam 1999). Jaron Lanier came to VR after a career in music, and introduced his company VPL's pioneering "Eye-phone" head-mounted displays as "the visual equivalent of audio headphones" (Jones 1993, 239). In more recent years John Carmack, until recently a leading engineer of the current VR boom as CTO of Oculus, continues to describe the VR head-mounted display as "headphones for your eyes" (Carmack 2018). Long before VR, audio engineers were fixated on producing the "illusion of reality" and providing the experience of "really being there" (Jones 1993, 241). But like VR today, the path to a broader cultural acceptance of headphone use was a rocky one. [43] Sketching out the long arc towards a broader social acceptance of headphone listening in the United States, Germany, and especially Japan, this essay traces the tumultuous history of head-mounted audio to demonstrate how virtual space first achieved cultural acceptance in the acoustic realm.

As with 3D visuals, the history of 3D audio can be understood as a series of attempts to replace a user's immediate perceptual surroundings with more malleable media alternatives. As Emily Thompson (2002, 7) has traced, by the first decades of the twentieth century "good sound" came to be associated with dry or space-less sound. Modern acoustics favored sounds stripped of their spatiality, and thus deliverable to audiences in forms more flexible, mobile, and self-

contained. Jonathan Sterne (2015, 111) describes this separation of spatial reverb from recorded sound as the emergence of the *detachable echo*. As he notes, it is now common for a single pop song to mix together many different acoustic spaces. For example, a vocal track may sound like it is recorded up close, while the accompanying guitars are drenched in reverb and sound more distant, with each occupying distinct acoustic spaces. From an early twentieth century perspective, this kind of bricolage might be expected to result in a perceptually fragmented experience, disorienting the listener by mixing different acoustic spaces together. Yet contemporary listeners are not the least bit confused.

Sterne explains this through reference to what he calls *speaker culture*, or how contemporary listeners are accustomed to hearing how sounds coming out of speakers (each with their own spatiality) blend with the immediate acoustics of their listening environment. He notes that if we define augmented reality (AR) as a process of layering virtual spaces on top of existing ones, speaker culture has produced augmented realities for a long time now. Sterne (119-20) uses this observation to caution against exaggerated claims that AR represents a radical transformation in human subjectivity, as humans immersed in speaker culture appear to have already adjusted to a world of layered virtual spaces just fine.

Yet by focusing on speakers as his archetypal case for the experience of "space within space," Sterne fails to note the more profound challenge presented by detachable echo when situated at closer proximity to the listener, as in the case of headphone use. Sterne's speaker culture can be understood as a variant of theatrical immersion, where both media playback device and listeners are situated within a shared physical space. An air gap remains between user and sound reproduction technology, allowing the acoustic qualities of the immediate physical environment to condition whatever sounds emerge from the speakers. Whereas Sterne proposes

contemporary listeners have simply become adept at mixing together different sonic spaces, the final acoustic mediation of the immediate physical environment still plays a crucial integrating role, grounding the various media layers through the more uniform and invariant acoustics of the surrounding space. No matter how complex the layering of virtual space on the recording, everything arrives sutured together by the reassuringly familiar spatial atmosphere of the surrounding environment.

With what we might call *headphone culture*, however, what becomes detachable is not just the acoustics of the original recording, but even this more basic spatial grounding of sound in the immediate physical environment. Isolating left and right ears by placing a speaker directly on top of each, headphones can reproduce a virtual acoustic space without filtering it through an existing physical location. Like the head-mounted [44] interfaces of VR, the virtual space of headphone sound seems to perceptually immerse a perceiver's body, but in fact only coalesces within their head. In other words, with headphone listening the spatiality of the listening situation is transferred entirely into the apparatus, becoming detachable *even from the listening space itself*. Not just a detachable echo, but a detachment of sound's role in embodied spatial awareness —what I will call a *detachable ambience*.

The detachable ambience of headphone sound has gone largely unstudied in existing sound studies scholarship on headphones, even as it emphasizes the emergence of new spaces for listening. Studies of the social aspects of headphone use have overwhelmingly focused on post-Walkman mobile listening, with an emphasis on personal media's ability to renegotiate the boundary between public and private space. In an early essay, Judith Williamson (1986, 210) notes how the Walkman "turns the inside of the head into a mobile home." In the mid-2000s, Michael Bull (2004, 5 and 33) describes how Apple's iPod headphones allow listeners to become

“enclosed” in “pleasurable and privatized sound bubbles,” while the white cables signal a desire to be left alone.<sup>1</sup> More recently, Mack Hagood (2019, 195) examines how Bose’s noise cancelling headphones allow users “to construct an on/off interface with the aural environment and the space one shares with others.”<sup>2</sup> While differing in whether they understand this privatization as empowering or alienating, these studies each situate headphone listening as part of a broader trajectory towards privatized media consumption, promising individual control over a personalized sound environment.

Approaches to headphone culture have remained up until now essentially *monophonic*, neglecting to consider the role of stereo and binaural sound spatialization in the history of headphone listening. Often drawing on Henri Lefebvre’s approach in *Rhythmanalysis* ([1992] 2004), much of this work focuses on how headphone use promises listeners the ability to transform their emotional or energetic states via the rhythmic resources of portable, private music. Despite frequent references to urban space, these studies situate the social role of sound recordings as primarily *temporal*: musically organized time intersecting, intensifying, or flattening out the broader rhythms of contemporary capitalism.<sup>3</sup> With the notable exception of Barry Blesser and Linda-Ruth Salter’s *Spaces Speak, Are You Listening?* (2006), sound studies scholarship is nearly silent on the [45] spatial acoustics of headphone sound, and only rarely attends to the architectural acoustics of physical spaces where headphone listening occurs. This spatially flattened approach has led sound scholars to overlook the historical importance of

---

<sup>1</sup> For a more recent example, see Mannering (2015).

<sup>2</sup> See also Thibaud (2003).

<sup>3</sup> For example, Ian Chambers (1990, 3) describes the Walkman as part of a transformation of the city from an “organizer of space” to a technological focus on the “organization of time.” While less explicitly about headphones, see also DeNora (2000), Fink (2005), LaBelle (2010), Roquet (2016a).

debates over stereo and binaural recording in the broader history of headphone culture. In this article, I aim to move beyond a simple “on/off” relationship between the sonic space *in* the headphones and the sonic space *around* the headphones, revealing a more complex historical intersection between the architectural acoustics of the built environment and the increasingly virtual ambient spatialization practices of headphone listening.

In recent years, 3D audio has once again emerged as a central focus for new headphone technologies.<sup>4</sup> The emergence of new consumer VR, AR, and so-called “spatial audio” technologies should push sound studies to grapple more directly with the theoretical and historical ramifications of mediated three-dimensional sound. As the intersecting history of headphone listening and virtual audio spatialization makes clear, the current cultural reception of VR and AR head-mounted displays — along with their attached headphones or speakers — was preceded and shaped by long-running debates over the spatial acoustics of headphone use. These debates largely played out in the pages of audio engineer and audiophile publications in countries like the United States, Germany, and Japan, only to spill over into more mainstream cultural contexts as headphone listening became popular globally in the post-Walkman era. While I draw on scholarship focused on the US and Germany in what follows, I emphasize Japan’s crucial but often neglected role in spatial audio history. The European and North American studies mentioned above often frame their work as revealing a specifically “Western” trajectory, yet the

---

<sup>4</sup> Bose’s wearable audio-only augmented reality glasses are the most prominent product along these lines so far, but spatial audio has made inroads in many areas of audio production and consumption. The Audio Engineering Society, the main professional organization for audio engineers, focused on mobile spatial audio and AR audio (along with assistive listening) as key themes for their 2019 conference. On sound AR, see Hagood (2018).

Japanese discourse reveals a great deal of transnational continuity (even as specific developments occurred at slightly different times in each place).<sup>5</sup>

While hard to imagine today, headphone use was frequently stigmatized throughout much of the twentieth century. Much like VR headsets, early headphone use was associated with spatial disorientation, physical discomfort, and feelings of isolation, even as it was shaped by broader social histories surrounding the intersection of media with things like gender and disability. In what follows, I propose it was only once the built environments of everyday life *also* became organized around detachable ambience that the headphone stigma began to fall away.

### **Speakers Attached to the Ears**

From their invention in the late nineteenth century, headphones were quickly understood to bring a person elsewhere, removing the listener from their immediate sensory environment. Early headphone models were used primarily as specialist equipment in telecommunications contexts (Yamamoto 1975, 225). Early switchboard operators at Bell, for example, used a workplace model complete with shoulder supports to help prop up the over six pound headset during long shifts. Telephone service was the primarily context for encountering headphones in Japan as well up through the early twentieth century, as in scenes featuring Murai Ayako wearing headphones while working as a switchboard operator in Mizoguchi Kenji's *Osaka Elegy* (*Naniwa erejī*, 1936).<sup>6</sup> Like phone booths more generally, the switchboard operator's headphones (and the

---

<sup>5</sup> See Bull (2004) for an example of the "Western" framing. This is despite important early Walkman studies focused on East Asian contexts including Hosokawa ([1984] 2012) and Chow (1990).

<sup>6</sup> On telephone switchboards in Japan, see Yoshimi ([1995] 2012), Yasar (2018, 48-51).



larger enclosure of the operator's booth) served as perceptual enclosures isolating the user from their immediate surroundings. As Sterne (2003, 158) writes of this early headphone context, "not only was hearing to be separated from the proximal auditory environment, but the act of communication itself was to be separated from the surrounding physical environment."

In one alternate history of cinematic immersion, audiences could have been listening to Mizoguchi's film *on* headphones as well. In the early days of Hollywood's conversion to sound film in the 1920s and 1930s, engineers at Bell Telephone Labs developed a sound system for cinema halls built around binaural audio provided through headphones, what Steve J. Wurtzler (2008, 263) calls the "logical outcome of developing sound technology on the model of the human body." Bell Labs is where the idea of positioning microphones on "dummy heads" to more accurately capture auditory perspective first emerged. Rather than strip away the acoustic detail of the original recording space, these recordings aimed to capture it more accurately by incorporating a model of the would-be listener's head for incoming sounds to reflect against. The dummy head technique attaches two monaural microphones to each side of a mannequin (usually just a head or head and upper torso) built to the dimensions of an "average" human body.<sup>7</sup> When recordings made in this way are later played back on headphones placed over the listener's ears (in approximately the [46] same position as the original microphones), the listener's brain registers the subtle ambient distortions produced by the dummy head as if it were being produced by their own body instead, enabling an enhanced sense of where the recorded sounds are situated in relation to the listener's body in the (now virtual) space of the recording. Researchers at

---

<sup>7</sup> Bell first used a dummy head named Oscar to produce a binaural recording of the Philadelphia Orchestra. Microphones at the time were too large to be situated within the dummy head's ears, so Oscar had microphones on his cheeks instead (Krebs 2017, 119-20). As with the examples below, the implied "average" listener here is explicitly gendered male.

Philips' research laboratories in Eindhoven, who had developed a similar artificial head (what they called a *Kunstkopf*) recording technique by the end of the 1930s, noted that sounds recorded in such a manner were "so real that listeners had to turn their heads to reassure themselves that nobody was standing behind" (quoted in Krebs 2017, 120).

Yet when Bell presented binaural headphone listening to movie theaters as a means to achieve better sound localization in films, the idea was quickly rejected in favor of more standard loudspeaker systems. While partly due to the cost of equipping each cinema seat with its own pair of headphones, theater owners were also skeptical audiences would be willing to perceptually isolate themselves in such a way. As Wurtzler (2008, 263) writes in his history of the era,

Such a binaural acoustic system would have posed a drastic contrast to preexisting theatrical filmgoing experiences. Would audiences watch and listen to a live theatrical presentation and then at some cue from the theater reach beneath their seats and put on their telephone headsets? Would they sing along, following an on-screen bouncing ball, aurally isolated by their personal acoustic reproduction device, not even able to hear the sound of their own voice? And what of the construction of the cinema as a live, group-based social experience?

Headphones would ultimately find a home in the cinema hall, but only for one very specific demographic: the hearing-impaired (Krebs 2017, 360fn81). The subsequent long-term association between disability and headphone use would underwrite, in part, the social stigma associated with headphones running all the way up until the 1970s. The isolation of headphone listening was figured as an option of last resort, reserved for the minority of audience members with hearing impairments. The privileged immediacy of not needing headphones in the theater became the default position, masking how the "speaker culture" of the movie theater itself was already built with these normative hearing bodies in mind.

Headphones for use with home record players were also on offer in the 1920s and 30s, but raised similar concerns about their isolating influence on the domestic scene. In the private sphere, headphone listeners at this time were also often the male 'head of household,' asserting social control over domestic space and trying (not always successfully, judging by the audio magazine commentary) to bend the surrounding acoustics to their will and to their hobby. Headphone stigma in this context often centered on the rest of the family's feelings of being left out. The editors of the *Living Stereo* anthology note how home headphone use in the US at this time was often perceived "to silence conversation in the home and exclude others from the radio experience," even giving rise to gendered figures like the "wireless widow." One 1924 letter to the editor of *Radio Broadcast* describes the struggle of one "wireless widow" forced to "sit mum" while her husband was occupied by his headphones, "consoling herself with a book, or solitaire, having at least the chill comfort of his physical presence, though his soul go marching on" (quoted in Théberge, Devine, and Everett 2015, 23-4). Some attempted to offset this isolation by creating devices for simultaneous group headphone listening. As Sterne [47] (2003, 161-7) has highlighted, home listening in the late nineteenth and early twentieth century offered accessories for allowing multiple family members to listen "alone together" to the same recording simultaneously through separate headphones. Yet much like group headphone use in the theater, it never really caught on.

When binaural reemerged in the postwar 1950s, it was a loudspeaker-driven phenomenon. Emory Cook, an audio engineer from New Jersey, kicked off the new binaural boom in 1952 with his invention of a two-channel stereo record system. Rather than printing a two-channel signal in a single groove (technologically impossible for the consumer records of the era), Cook's system used a special two-pronged record arm to simultaneously pick up sounds

from two grooves at once. Echoing the 3D film boom occurring at the same time, Cook promoted his new system widely as an innovation in “three-dimensional sound,” presenting his own field recordings of passing trains and other moving sound sources in ways that emphasized spatial depth.

A number of public events in Japan this same year introduced the Japanese public to binaural speaker broadcasts based on Cook’s system (Iwai 1955, 141). These included a 3D sound broadcast using speakers on multiple busses set up in Hibiya Park in Tokyo, as well as indoor demonstrations at the International Radio Center. Starting in late 1953 national broadcaster NHK launched a special binaural radio broadcast once a month where the left binaural channel was broadcast on one frequency, and the right channel on another. The broadcasts were a popular success, and the system had the advantage of remaining audible to owners of only one radio (who could listen in mono), while at the same time promoting the purchase (or at least the borrowing) of a second radio set to enable the full binaural experience.<sup>8</sup> Writing in 1955, Tokyo Television sound researcher Iwai Seiji called binaural recording a “revolution in the audio world,” but noted many technical issues remained to be solved. While the engineering aspiration was still towards better sound reproduction for music playback, binaural researchers were clear 3D binaural spatialization lent itself best to radio dramas and sound effects (like the steam trains and ping pong sounds on Cook’s records) (Iwai 1955, 140-1).<sup>9</sup>

---

<sup>8</sup> Cook’s binaural records were also available as imports for purchase, but were still too expensive for average Japanese consumers at the time (priced around 3,000 yen or \$89 USD in 1955 prices). As *New Electronics* [*Shin denki*] magazine complained at the time (1955, 113), to experience 3D sound it was still far cheaper to just go to a live concert.

<sup>9</sup> See also *Sutereo haifai seisaku dokuhon* (1962, 14).

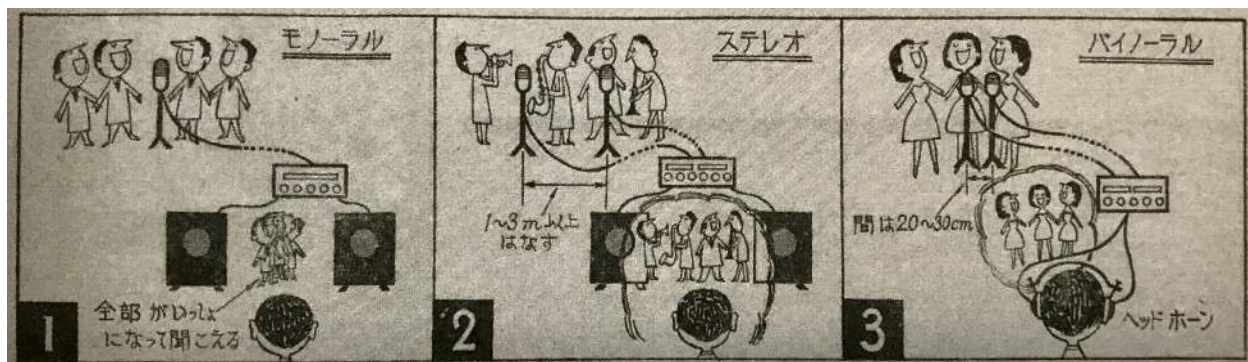
At the same time, proponents of listening to binaural recordings on headphones in the immediate postwar decades continued to face the stigma attached to headphone use. Audio engineers like Cook acknowledged the superior “fidelity” and “realism” of binaural headphone listening, but argued the available “earphones” were “both uncomfortable and anti-social.” Promoting 3D sound through loudspeakers instead, he believed, could turn binaural “from a tinkerer’s hobby into a potentially popular medium” (quoted in Théberge, Devine, and Everett 2015, 24). Through the mid-1970s, headphones remained limited either to bulky high-end headsets for audiophiles, or simple monaural earbuds with poor sound quality. A 1976 article in the Japanese *Radio Technology* [*Rajio gijutsu*] magazine (Nishimaki 1976, 248) notes that headphones capable of better quality sound reproduction tended to be heavy, meaning they were difficult to use for long periods and could even make the listener sweat in summer. Smaller earbud-type headphones, meanwhile, could not achieve good enough audio quality for music listening. As with cinema headphones, the latter were by this point popularly associated primarily with hearing aids, adding another dimension to the ableist headphone stigma (Weber 2010, 353).<sup>10</sup>

However, researchers and audiophiles in Japan and West Germany were reluctant to give up on headphone listening entirely. Japanese hi-fi enthusiast magazines in the 1950s and 60s featured numerous articles breaking down the relative merits of different [48] approaches to spatialized sound recording, as well as speaker versus headphone playback. For example, the *Guidebook to Building a Stereo Hi-fi* [*Sutereo haifai seisaku dokuhon*], released in the midst of the Japanese consumer hi-fi boom of 1962, sets out to illustrate the different effects of monaural, stereo, and binaural listening for Japanese audio enthusiasts. As depicted in the accompanying

---

<sup>10</sup> On the long-running quest to render hearing aids ‘invisible,’ see Mills (2011).

illustration (Figure 1), the authors note how single-channel monaural recording creates a sound stage where the listener perceives all the sounds as if clumped together in one spot. Recording in stereo with two microphones at least 1-3 meters apart creates a two dimensional sound field, with the different performers sounding as if they were arrayed in a line in between the two speakers. Binaural recording, in contrast, uses two microphones placed only 20-30 centimeters apart (based on an inter-ear distance of twenty-three centimeters for the “average” human). When listened to on headphones, a recording produced in this way creates a more intimate three-dimensional sound stage in close proximity to the listener’s head (*Sutereo haifai seisaku dokuhon* [1962], 14).



**Figure 1. Illustration from *Guidebook to Building a Stereo Hi-fi* (*Sutereo haifai seisaku dokuhon* [1962], 14) showing the perceptual difference between audio localization in monaural (left), stereo (center), and binaural (right) recording and playback.**

While the text doesn’t mention it, the *Guidebook to Building a Stereo Hi-Fi* illustrator switches from the all-male quartets portrayed in the first two panels to a female vocal trio for the binaural listening illustration. For the implied male listener (seen in rear profile here), the greater intimacy of binaural sound appears to bring with it a different set of gendered expectations.

Stefan Krebs (2017, 139) notes how the primarily male recordists, trade journalists, and radio professionals in the 1970s German recording world readily assumed a male head would “represent an average listener,” and the situation was very similar in Japan. As evidenced by the “wireless widows” of the 1920s, domestic control over home listening technology has also long been assumed to be a primarily a male domain.<sup>11</sup> Yet the gendering of binaural *performance* skews heavily female.

Notice the shift in representational strategies when binaural headphone listening brings the soundstage up close. Rather than bringing other men inside the male listener’s head, the premise shifts to a promise of close proximity with (virtual) women. This gender switch aligns with the preponderance of female voices and vocalists on Japanese demonstration records for binaural sound from the 1950s through the 1970s. For example, Teichiku’s *Headphonic Binaural Record*, released in coordination with Matsushita in 1976 as part of the research developments detailed below, includes “narration” by voice actress Suzuki Hiroko, and ends with her “Love Talk” [*ai no katarai*] whispered into the listener’s ears. The switch to an ostensibly less threatening gender (given the gender politics of the [49] time) who performs affection for the implied male listener attempts to short-circuit what otherwise might be felt as an uncomfortable degree of intimacy and immediacy manifested in binaural listening — perhaps even a challenge to the spatial authority of the male home listener.<sup>12</sup>

As prices dropped and incomes rose, stereo record players for home listening became a key part of the Japanese “leisure boom” of the early 1960s (Wakabayashi 1961, 162).

---

<sup>11</sup> See also Keightley (1996) and Weber (2010, 347).

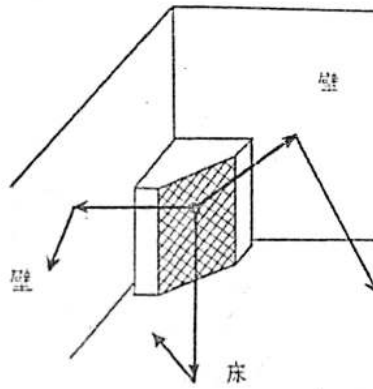
<sup>12</sup> The skewed gender ratio of producers and engineers being overwhelmingly male but characters and persona populating virtual spaces presenting as predominantly female persists in the context of Japanese VR today. On the gendering of proximate binaural whispers in a more recent ASMR context, see Andersen (2015).

Headphones [*heddo-hon* or *heddo-phon*, also called earphones/*iyā-phon*] for consumer use emerged as part of the boom, particularly for use with transistor radios, but as noted above they were not without their detractors. Writing in *Record Art* [*Rekōdo geijutsu*] magazine in 1961, Wakabayashi Shunsuke (162) differentiates the virtual acoustics of headphone listening for his readers like this:

In the case of listening to stereo sounds through earphones, sounds go directly into the ears. There is no mediation. In contrast, when listening to stereo recordings on speakers, sounds pass through the medium of air [*kūki*] before arriving at the ears. Because sounds coming from speakers pass briefly through the room before arriving at the ears, the sound includes reverberation off the surrounding walls. Depending on the power, sounds from the left speaker may also be audible to the right ear. Compared with sound passing through space like this, with headphone use sounds instead reverberate directly from the outer ear into the eardrum. There is a big difference in sound.

For Wakabayashi, speaker listening bounces the recorded music off the reflective surfaces of the listening space (walls, floor, and windows, for example), producing "stereo sound's unique feeling of presence" [*sutereo dokutoku no rinjōkan*] (Figure 2). As I noted above, [50] headphone listening in contrast effectively bypasses the local ambience of the listening space by collapsing the distance between ears and speakers, isolating the sounds coming into each ear from each other and from the larger listening environment.





**Figure 2. Illustration from Wakabayashi (1961). Caption: “Sounds coming out of the speaker reflect off walls, the floor, and other places before reaching the ears.”**

Wakabayashi cautions that if recordings designed for stereo speakers are listened to with headphones, some unfortunate effects occur. For example, if a standard stereo recording of an orchestra is listened to over speakers, the listener is acoustically positioned as if they were in the concert hall audience. If the same recording is listened to over headphones, however, the “sound image” suddenly comes much closer, as if the listener is positioned right in the middle of the orchestra pit. For sound recordings produced with speaker playback in mind, listening on headphones tends to give the impression the sounds are simply moving around inside the listener’s head (Wakabayashi 1961, 162-3). For this reason, Wakabayashi notes, headphone listening is “not well suited to real music appreciation.” Noting how many record producers at the time strongly advise against using headphones to listen to their recordings, Wakabayashi concludes that speaker listening generally provides a better experience, particularly in the case of music. Echoing Iwai’s earlier comments from the 1950s, Wakabayashi argues headphone listening is the best choice only for binaurally reproducing “real sounds” [*genjitsu-on*] like steam

trains and jet engines, or for more accurately localizing discreet sound sources like opera soloists or actors in a radio drama.

Revealingly, however, Wakabayashi's article ends by moving away from these concerns with pure audio fidelity, pivoting to a separate set of issues that would prove decisive for headphone use in the coming decades. Here he admits the real appeal of headphone listening may lie elsewhere:

Despite all this, when thinking about the current situation in Japan, it's true most ordinary Japanese rooms are not very large, and moreover insulation between neighboring rooms is bad, so it is easy to hear sounds from next door. From this perspective, we can say that earphones offer a really easy way to listen to stereo recordings. It's difficult to overlook the appeal of being able to enjoy recordings on your own, without bothering other people (163).

Wakabayashi explicitly warns readers that from a fidelity perspective, listening to stereo recordings on headphones has the "extremely dangerous" [*kiwamete kiken na*] possibility of producing a poor balance between the different instruments, and suggests listeners compare for themselves by using the same recording for both speaker and headphone playback. Yet he admits "late night stereo listening on headphones by yourself can also be fun" (163). The appeal of headphones here is ultimately less about sound quality and more about the provision of a sound environment scaled to a single individual. Heike Weber (2010, 343-6) notes a similar emphasis on "listening without bothering anyone" in West German headphone advertising at the time. As with Japan, this was in part due to the poor sound insulation in German housing.

As Wakabayashi's comments reveal, by the 1960s headphones' ability to provide a solitary listening space was beginning to register not as a liability, but as a primary part of their appeal. This was also the decade headphone use for film audiences finally found popular acceptance, albeit only in a very specific context: the in-flight movie. In the airplane cabin, the

alienating potential of headphones was transformed into an asset for the solo traveler. The headphone enclosure offered a perceptual escape not just from plane noise but from the discomforts of the newest air travel innovation: the reduction in [51] personal space known as Economy Class seating. In-flight headphones enabled what Stephen Groening (2016, 117) calls the "separated spectator" in the skies. Even as passengers were seated ever closer together, the virtual space provided by the headphones ensured they were perceptually apart.

As Wakabayashi's article hints, headphone use at home was in the midst of a similar, if more staggered transformation. Headphone listening offered a compelling chance to bypass existing home acoustics by piping the virtual space of a recording directly into a listener's ears. But the sound perspective issues persisted. If recordings were designed for speakers and thus recorded with microphones placed several meters apart, this larger sound stage would effectively be compressed into the smaller space between the ears when played back on headphones.<sup>13</sup> This resulted in what came to be known around this time as the "sound in the head" problem, illustrated with drawings showing what it feels like to listen to stereo recordings on headphones. For example, a stereo recording of piano music would sound as if you have placed your head directly inside the performer's piano (Figure 3).

---

<sup>13</sup> See Blesser and Salter (2006, 186-91) for a detailed examination of the acoustics of in-head localization.



**Figure 3. Illustration from *Radio Technology [Rajio gijutsu]* magazine (July 1976) illustrating the problem with listening to stereo recordings on regular headphones. Caption: “Sounds without reflection = it’s as if you are listening having put your head inside the instrument” (Yamada and Kikuchi 1976, 251).**

### **Getting the Sound Back Outside the Head**

In order to provide an appropriately-sized virtual soundstage for headphone listeners, the sound somehow needed to be moved back out into the space surrounding the body. Attempts to solve the sound in the head problem would drive an increasing separation between “binaural” and “stereo” recordings in the decades to follow. While Cook’s binaural records were intended for speakers and recorded with a pair of microphones placed several meters apart, the new binaural recording practices instead took a cue from Bell’s earlier “dummy head” approach, spacing the microphones roughly the width of a human head.

[52] A breakthrough came in 1972, with an updated artificial head microphone (*Kunstkopf*) designed by Berlin-based researchers Georg Plenge, Ralf Kürer, and Henning Wilkens. Microphones could be made small enough by this point to place them inside the ear canal of an artificial head, making it possible to capture the acoustic perspective provided by the

outer ears as well. This new form of binaural was first introduced to the West German public in 1973 at the International Broadcasting Fair (IFA) in Berlin, followed by the broadcast of the first German binaural radio drama. This kicked off a small boom in binaural headphone listening in West Germany that lasted through the end of the decade, at which point most home audio equipment would include headphone jacks as a standard feature (Krebs 2017, 117).

As Stefan Krebs notes, while many daily newspapers in West Germany were enthusiastic about the new three-dimensional sound, and binaural was popular with listeners, the audio engineer trade journals were less enthusiastic. As with the Japanese engineers, a range of technical and infrastructural compatibility concerns emerged, such as how binaural recordings were effective in locating sounds on a left-right horizontal axis, but had much more difficulty positioning sounds in front or in back of the listener.<sup>14</sup> Krebs (123) notes headphones at the time were still "associated with the early years of radio broadcasting, and thus their use was perceived as a step backwards in technical terms." Making headphone use mandatory for listening to binaural broadcasts also remained a sticking point. As one trade author asks rhetorically in 1973, "who is willing to submit themselves to these restrictions?" (quoted in Krebs 2017, 124). Unable to secure larger institutional support, the 1970s West German binaural boom faded by the end of the decade.

In a striking parallel with the Japanese sound magazine discourse from a decade prior, binaural headphone experiences even led some German audio experts to question the implicit

---

<sup>14</sup> As Blesser and Salter note (2006, 188) this problem occurs because individual differences in the shape of the external ears (the pinna) are crucial to how the brain locates sounds in front or behind the body – a generic or averaged outer ear like the dummy head's is ineffective. Recent binaural research has focused on how to incorporate individual 3D scans of a listener's pinna into the spatial audio processing to provide a more custom virtual localization. In the Japan context see Nishimaki (1976, 250).

assumption that greater three-dimensional access to "objective reality" would inherently lead to better or more satisfying listening experiences. For example, Peter Reinecke, director of the Berlin Institute of Musicology, declared at a 1969 conference the "need to question the ideological position of 'high fidelity,'" because "music is not sound." As Wilhelm Schlemm of Radio Free Berlin described in 1974, "Kunstkopf stereophony has come at an untimely moment, leading to the following paradox: previously the history of electro-acoustical transmission was always driven by the desire for highest fidelity [...]. [N]ow, just as the aim of perfect illusion can be achieved, it is no longer valued and desired" (both quotes in Krebs 2017, 133-4).<sup>15</sup> Consumer demand for flexibility in listening arrangements was overtaking the desire for accurate sound reproduction in and of itself, even as binaural engineering remain fixated on the latter.

To complicate the situation further, as Krebs notes, the idea of perfect binaural fidelity to the original recording space also seemed to threaten the role of the audio engineers themselves, who in the new era of post-production spatial control had come to see their work as an active and integral part of the larger creative process of producing musical recordings. While the quest for increased fidelity and realism had long led audio engineering rhetoric, in practice the sculpting of acoustic space had already moved far away from the original recording environment, placed under the control of both studio engineers and home listeners instead. Having once grown accustomed to controlling the acoustics virtually, both parties were not about to sacrifice this flexibility in deference to the fixed "realism" of the original recording space.

[53] Similar debates over headphone listening would captivate Japanese audio hobbyists in the mid-1970s. Yamamoto Shūhei observes in a 1975 article for *Radio Technology* [*Rajio gijitsu*] that headphone listening had finally emerged as a "quiet boom" [*shizuka na būmu*]

---

<sup>15</sup> On the complex history of promises of audio fidelity, see Sterne (2003, 215-86).

among audio enthusiasts, yet the sound in the head problem, the weight and discomfort of high-quality headphones, and the market dominance of stereo recordings intended for speaker playback all remained as obstacles.

Japanese audio engineers threw themselves at these problems in the years to follow, hoping to create an acoustic architecture that could retain the privacy and convenience of headphone use while “more naturally incorporating the atmosphere of the surrounding room” (Yamamoto 1975, 225-6). Early experiments included a head-mounted speaker system designed to project a stereo recording to sound as if it were emerging from directly in front of where the listener was facing — a kind of audio-only head-mounted display (Figure 4).

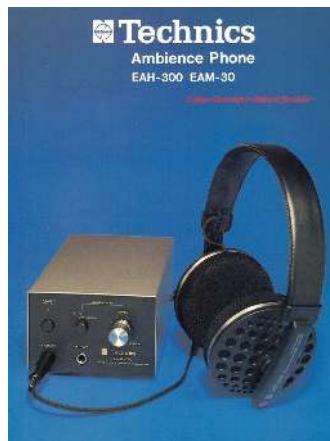


**Figure 4. Caption: “Headphones that can position sound in front of the head” (Okahara 1977, 112).**

Further developments along this line came from the Acoustic Research Laboratory at Matsushita, which came up with a different approach to what they called “outer head localization” (Gotō et al. 1976; Gotō, Kimura, and Watanabe 1976). This research ultimately resulted in a consumer product released in 1977 on their Technics label: the Ambience Phone

[*anbiensu-hon*]. Explicitly marketed as a solution to the “sound in the head” problem, the Ambience Phone used four-channel audio processing to mix additional artificial echo and reverb into any stereo recording sent through the device. Instead of hearing sounds emerge from the space between the ears, listeners would hear them coming from a pair of (virtual) speakers situated somewhere across the room (Yamada and Kikuchi 1976, 251). The Ambience Phone sought to approximate the intended acoustics of Sterne’s stereo-based “speaker culture” within an individual head-mounted space.

[54] The Ambience Phone has just three controls: an on/off button, a button to turn the ambience function on and off, and a dial allowing the user to set the desired amount of ambience to add to the original signal (Figure 5). By shifting the choice of stereo-vs-binaural sound placement from the producer to the consumer, the device also sought to present a workaround to the distribution and infrastructure issues binaural had repeatedly faced. By utilizing the on-board “ambient controller,” every listener was in effect put in charge of their very own (virtual) dummy head, turning them on or off with the push of a button and moving them forward and back in virtual space with the turn of a dial. Moving beyond the detachable echo of the recording studio, the Ambience Phone aimed to put detachable ambience under the control of the home listener.





**Figure 5. The Technics Ambience Phone (1977).**

The Ambience Phone was just one of a range of new forms of detachable ambience emerging around this time. One year later, Amar Bose would come up with the idea for noise-cancelling headphones while commuting by plane (Groening 2016, 133). Brian Eno coined the term "ambient music" around the same time to describe a genre aiming to provide self-selected atmospheres for private listeners, often by providing cavernous reverb-driven spaces within the recordings themselves.<sup>16</sup> And in 1979, Sony released the Walkman, a device that didn't solve the "sound in the head" problem directly, but effectively displaced it by opening up new possibilities for mobile headphone listening.

Histories of headphone use generally point to the Walkman's combination of lightweight over-the-ear headphones with a portable cassette stereo as the moment when [55] headphone use finally goes mainstream. Yet this precise combination of perceptual affordances was far from a foregone conclusion. Independent research groups at Sony had been working separately on lightweight headphones and portable cassette players. The cassette player group had originally planned on releasing the device with a pair of much larger, heavier headphones, assuming sound quality would be the most important criteria for their customers. It was already late in the development process when Sony co-founder Ibuka Masaru recommended the two products be combined.<sup>17</sup> The lightweight foam-covered headphones that were eventually used employed a patent Sony had licensed from Sennheiser, but miniaturized the design even further (Weber

---

<sup>16</sup> See Roquet (2016a) for more on the broader turn to ambient aesthetics at the time.

<sup>17</sup> See Shu Ueyama excerpt in Gay et. al. (2013, 131-2).

2010, 352). Ibuka was prescient in realizing that for most consumers, better fidelity was no match for convenience and comfort.<sup>18</sup>

The second unexpected development for the Walkman occurred after the first version of the device was already released to the public. Sony president Morita Akio had initially believed, based on his wife’s input, that listening alone in public would be considered rude. For this reason, a second headphone jack was included with the first release, much like the multi-headphone setups of home record players in the 1920s and 30s. Early Walkman ads emphasized how the device could be shared, for example, by a “very tall American woman and an old Japanese gentleman,” or two lovers on a tandem bike (quoted in Gay et al. 2013, 59). While later Walkman studies often present this as a historical curiosity, Morita’s wife’s concern was grounded in decades of stigma against the isolation of solitary headphone use, including the gendered “wireless widow” phenomenon of a half-century earlier.

Yet after the Walkman was put on the market, it became clear listeners were using the device primarily for listening alone. As Morita later wrote, “buyers began to see their little portable stereo sets as very personal” (quoted in Gay et al. 2013, 54). Shūhei Hosokawa later dramatized this isolation in his widely-read 1984 essay on the Walkman: “The listener seems to cut the auditory contact with the outer world where he really lives: seeking the perfection of his ‘individual’ zone of listening” (Hosokawa [1984] 2012, 106). To Sony’s surprise, listeners began treating the Walkman as a private extension of their own sound environment, enjoying the freedom it provided to ignore — or at least remediate — the surrounding perceptual world. While Walkman users too were subject to the headphone stigma early on (often seen as rude

---

<sup>18</sup> Facebook would later make a similar bet by focusing on less powerful but more convenient ‘stand-alone’ VR headsets like the Oculus Go and Oculus Quest.

young people purposefully refusing to attend to the larger social atmosphere), soon enough an increasing cultural acceptance of public headphone use emerged in Japan, across all age groups (Hosokawa [1984] 2012). Weber (2010, 353) notes a similar trend in the early Walkman reception in West Germany, as does Tuhus-Dubrow (2017, 44) in the US. Importantly, the Walkman also allowed for the isolating affordances of headphone listening to be more accessible across gender lines, as Alexander Weheliye (2005, 133) notes in his reading of the device as offering an "autonomous sonic space" for women who may not have control over space otherwise either at work or at home.

As noted above, scholarship on personal stereos has focused on how the listener's newfound mobility allowed private music to intersect with and reconfigure the embodied experience of moving through a city. What has yet to be recognized is how the Walkman normalized the practice of strapping stereo recordings directly to the head. Both the Ambience Phone and the Walkman sought to offset the acoustic isolation of headphone listening by virtually repositioning stereo recordings back out into the surrounding world. [56] The former used audio signal processing to fabricate a detachable ambience directly within the device, while the latter relied on live on-location sound to surround the mobile listener and blend with the stereo recording, distracting them from the "sound in the head" effect. In this way, the Ambience Phone and the Walkman offered a sonic prelude to virtual reality and augmented reality, respectively (minus the head-tracking audio technologies that would only arrive a decade later).<sup>19</sup>

---

<sup>19</sup> See Stankieveh (2007) for later sound art examples.

## The Emergence of the One Person Space

Just as new sound technologies were making ambient control more accessible to the individual listener, the built environment was itself becoming more oriented towards a privately-mediated ambience. Wakabayashi, remember, noted in the early 1960s that home headphone use was fun in part because it allowed for late-night listening where speaker playback would otherwise be bothersome [*meiwaku*] for neighbors and family members in adjoining rooms. Record numbers of Japanese were relocating to urban environments at this time, often living in wooden housing with notoriously thin walls and in close proximity to neighboring homes. While these noise concerns were not exactly a novel phenomenon — noisy neighbors were even the central conceit of Japan's first sound film, *The Neighbor's Wife and Mine* (*Madamu to nyōbō*, Goshō Heinosuke, 1931) — the audibility of the neighborhood clearly impinged upon the promised freedoms of the 1960s "leisure boom," including being able to make full use of your home hi-fi as you saw fit.

As Sterne traces in *The Audible Past* (2003, 160), from the early days of sound reproduction technologies emergent middle classes have often sought to claim individual acoustic spaces as a form of private property. For the rapidly expanding middle class of high-growth Japan this first meant a shift to American-style one-room-per-child family homes, an unheard-of level of personal space for most Japanese children at the time. Weber notes a similar trend in 1970s West Germany, where domestic media consumption was becoming increasingly atomized, more about pursuing individual media habits rather than collective family-oriented consumption. As in Japan, this began with teenagers in their bedrooms and soon extended to other family members as well. Headphone advertising subsequently shifted to promote an ideal

of isolated, concentrated listening, even associating it with relaxation and meditation practices (Weber 2010, 347-51).<sup>20</sup>

This trend was later taken to its logical extreme with the rise of concrete apartment blocks of one-room studios, or what in Japanese is called the “one room mansion” [*wan rūmu manshon*]. As Nango Yoshikazu describes in his *Urban Theory of the One Person Space* [*Hitori kūkan no toshiron*, 2018], when one room mansions first emerged in the mid-1970s these single-occupant dwellings were intended as weekday places to sleep for urban workers who would return to their families in the suburbs on weekends. But they soon began catering to the rising number of one person households willing to compromise on space in order to secure a space all of their own in the city. The one-room mansion was advertised as enabling what in Japanese is known as “my pace” [*mai pēsu*] — a life lived in your own way and at your own speed, even if it has to take place within rather tight quarters (89). A crucial element enabling this isolation was the extremely effective sound insulation, which made it almost impossible to hear what was going on in the unit next door (92).

[57] As Nango notes, while these micro-apartments were often derided as “rabbit hutches” by the American and European press, much like the Walkman they ended up catering to an unexpectedly wide range of ages, income levels, and occupations. Rather than bemoan the small space, Japanese media at the time recognized the one-room lifestyle as having its own distinct appeal. Photobooks like Tsuzuki Kyōichi’s *Tokyo Style* (1993) presented the one room mansion as a kind of “cockpit” [*kokkupitto*] for the solo dweller: occupants could lie in bed and still have everything necessary for controlling their media environment within ready reach (Nango 2018, 98). During the one room mansion’s heyday from 1975 to 1985, this personalized

---

<sup>20</sup> On similar trends in Japan, see Roquet (2016a); in the US, see Hagood (2019).

cockpit might have been furnished with a television, a stereo, and perhaps later in the decade a video cassette deck or video game console.<sup>21</sup>

Building on Okada Toshio's work, Thomas Lamarre (2018, 200-1) argues that around this time in Japan the private television set began to serve as a pivot for "switching between media functions," such that "the personal bedroom begins to look like a broadcast-relay station." Along with devices like the Walkman and Ambience Phone, the architectural acoustics of the one room mansion reveals a concurrent virtualization of private perceptual space occurring in tandem with the establishment of these new media networks. The solo acoustics of the one person cockpit reimagined perception along the lines of a VR flight simulator, blocking out the immediate surroundings in order to allow for greater immersion in virtual space. The arrival of this acoustically sealed yet virtually expansive solo media environment can be understood as a further normalization of the detachable ambience that had long been sought through headphone listening. Never simply about privacy, the new aural architecture aimed to open out a more expansive space of mediated perception organized around the body of a single individual.

### **Towards a Head-Trackled Acoustics**

While the urban environments of late twentieth century Japan were increasingly structured around an ideal of one person space, audio engineers continued trying to find ways to use virtual acoustics to transcend the immediate listening environment. Binaural reemerged in the post-

---

<sup>21</sup> A related phenomenon emerged in the US by the mid-1980s, popularized by marketing consultant Faith Popcorn as *cocooning*, a practice of staying inside the home surrounded by video cassette recorders and compact disc players rather than venturing out into the "scary" outside world. Popcorn describes cocooning in 1987 as "a rapidly accelerating trend toward insulating oneself from the harsh realities of the outside world and building the perfect environment to reflect one's personal needs and fantasies." Quoted in Krier (1987).

Walkman 1980s in a few scattered experiments in North America like Norman Durkee's binaural opera *Oxymora* (complete with headphones provided to the audience), and a binaural recording of Steven King's short story *The Mist* produced by ZBS radio in upstate New York (both 1984). A brief hype also emerged surrounding Argentinian inventor Hugo Zuccarelli's "holophonic" binaural system. In a 1983 article in *The New Scientist*, Zuccarelli claimed the human auditory system emitted a steady sound that could be used as a reference tone for enhanced audio spatialization, and sought to market the idea through a closely-guarded proprietary binaural technique he dubbed "holophonics" (riffing off the popularity of 3D holograms at the time). The system was deployed in early 1980s albums by Psychic TV, Paul McCartney, and Pink Floyd, and on demonstration albums focused on close-to-the-head sound effects, like a recording of scissors moving around the ears during a haircut. Takemura Mitsuhiro published two articles building anticipation for the technology in Japan around the time of the demonstration record's Japanese release in the summer of 1987, but even these articles note the system was shrouded in "mystification," with Zuccarelli exhibiting a suspicious reluctance [58] to let outsiders verify his claims (1987a, 165; 1987b, 21). In retrospect, it appears likely this secrecy was to avoid further scrutiny of the questionable science behind the system, as several responses to the *New Scientist* article had quickly pointed out.<sup>22</sup>

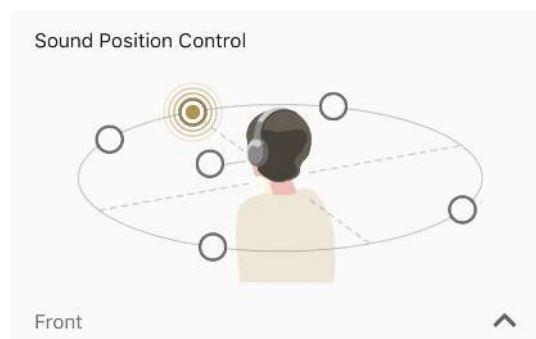
As Takemura notes (1987a, 163), even in the post-Walkman years the fact that binaural systems like Zuccarelli's demanded headphone use to achieve full 3D audio perspective remained an obstacle to their popular adoption. While mainstream acceptance remained elusive, however, 3D spatial sound using binaural methods was quietly making further inroads into audio

---

<sup>22</sup> For a helpful breakdown of binaural and holophonic techniques and issues with the latter, see Mark A. Jay's posts to the *Gearslutz* discussion board in the "holophonic vs. binaural" thread (Jay 2010).

production pipelines. For example, systems like the 1991 Roland Sound Space Processor promised audio engineers the ability to computationally position sounds within a virtual 360-degree sound field, with the system calculating the appropriate spatial acoustics given the source location and the specifications of the virtual environment.<sup>23</sup>

On the consumer side, the outer head localization of the Ambience Phone lives on as a feature in higher-end Japanese headphones, which today come with an ever-expanding set of ambient controls. For example, Sony’s current premier line of noise cancelling headphones, the WH-100XM series, includes a smartphone app with “Ambient Sound Control” options for listeners to select or have automatically chosen for them based on the headphones’ on-board sensors. The app includes settings for listening in place, while walking, running, and in transit; the option to apply various concert hall and arena acoustics; and, closest to the original Ambience Phone function, a set of “Sound Position Control” options for virtually displacing the auditory perspective of a stereo sound recording to one of five locations further away from the head (Figure 6).



**Figure 6. Sound Position Control settings in the iOS app for Sony’s WH-100XM2 headphones (Captured October 20, 2018).**

<sup>23</sup> See Jones (1993, 244), Blesser and Salter (2006, 194-98).



Each of these systems present an acoustic space positionally fixed in relation to the orientation of the headphone listener's head. If the person turns to face a different direction, the entire virtual soundstage pivots with them. In contrast, when binaural audio is combined with head-tracking technologies, the sound field can be updated in real time in precise alignment with the listener's movement. The first head-tracked binaural audio was realized in 1988 (Grau 2003, 168), and current VR and AR headsets allow auditory perspective to be mapped in real-time to the orientation of a listener's head (pivoting around a single point using a gyroscope, and possibly also tracking spatial [59] displacement of the head as well). For example, if a listener turns their head to the left, the virtual soundstage produced by the headset immediately pivots an equal and opposite degree to the right, giving the impression the individual sound sources have stayed in place while only the user's head has moved.

From the perspective of the sound processing system, the listener becomes just one more object to position within the virtual sound field. From the listener's perspective, the sound environment seems to persist independently in 3D virtual space, detaching the ambience not just from the physical environment but from the listener's own body as well. In effect, head-tracking allows listeners to more fully guide their own 'dummy head' through virtual space in real time, more fully replacing the acoustics of the immediate physical environment with the detachable ambience of a mediated virtual soundstage. Steven Jones (1993, 240) notes how the spatial persistence of virtual audio independent of a listener's movements plays a crucial role in VR immersion, offering a way to "further distract the user's attention from the nonvirtual world, to solidify the technology's 'authority over the observer.'"<sup>24</sup> In this way the headset becomes "a

---

<sup>24</sup> The quote is from Crary (1990, 102).

cockpit that the pilot wears,” as VR pioneer Thomas Furness once described it (1988, 5) — a one person space directly attached to the face.

While head tracking introduces its own specific issues, the history of headphone stigma and the “sound in the head” problem reveals many precedents helpful for understanding the current social tensions surrounding the newer head-mounted media formats, as well as where the technology might be headed. One example is the intersection of head-mounted media and public space. Writing in the mid-1980s in one of the earliest critical studies of the Walkman, Judith Williamson (1986, 211) predicted that “personalized mobile viewing would be the logical next step” for portable media, given visual media consumption had already, like sound, moved from the public theater to the private home. “Imagine people walking round the streets with little TVs strapped in front of their eyes,” she writes, essentially describing early VR head-mounted displays.<sup>25</sup> Sure enough, following the emergence of stand-alone consumer VR headsets people are already testing the cultural acceptability of VR in public. In Tokyo some early adopters of Facebook’s Oculus Go headset have ventured out in public with the device, riding on trains and gathering in restaurants to test how mobile VR might be deployed in public space much like headphones already are.<sup>26</sup> According to participants the response from passerby was largely positive. Yet someone also tried using a VR headset on a Boston Green Line train in 2016, only to be mocked on social media and criticized by the local police for making themselves a “soft

---

<sup>25</sup> The phrase was echoed in a recent viral hashtag on Japanese Twitter describing the VR headset as a “tv attached to your face” [*atama ni tsukeru terebi*], after Tokyo musician Katsunori Adachi noted his wife had used the phrase to search for a picture of the Oculus Go headset online. Some early hand-built VR headsets did, in fact, use a pair of small portable televisions for the head-mounted display.

<sup>26</sup> The “Densha de Go” train gathering was organized by the Oculus Go o tanoshimu kai [Society for enjoying the Oculus Go] Discord group and held on May 19, 2018.

target” for thieves (Annear 2016).<sup>27</sup> Here too the cultural acceptance of VR is playing out differently in different social contexts, though as headphone history reveals it may still shift dramatically over the long term. If the history of binaural audio offers any guide, the detachable ambience of virtual reality may only come to be more fully embraced once the surrounding environments of everyday life become more virtual themselves.

**Paul Roquet** is associate professor of Media Studies and Japan Studies at the Massachusetts Institute of Technology (Comparative Media Studies/Writing), and the author of *Ambient Media: Japanese Atmospheres of Self* (Minnesota, 2016). His work focuses on the cultural politics of mediated space in contemporary Japan. <https://proquet.mit.edu>

#### Funding Details

This work was supported by a Japan Foundation Japanese Studies research fellowship (long term).

#### **References**

- Andersen, Joceline. 2015. “Now You’ve Got the Shiveries: Affect, Intimacy, and the ASMR Whisper Community.” *Television & New Media* 16 (8): 683–700.
- Annear, Steve. 2016. “Please Keep Your Virtual Reality Off the T.” *Boston Globe*, April 1.
- Blessner, Berry and Linda-Ruth Salter. 2006. *Spaces Speak, Are You Listening?: Experiencing Aural Architecture*. Cambridge, MA: MIT Press.
- Bull, Michael. 2004. “Thinking about Sound, Proximity, and Distance in Western Experience: The Case of Odysseus’s Walkman.” In *Hearing Cultures: Essays on Sound, Listening and Modernity*. Viet Erlmann, ed., 173-91. Oxford: Berg.
- . 2007. *Sound Moves: iPod Culture and Urban Experience*. London: Routledge.

---

<sup>27</sup> Augmented reality devices have also triggered social pushback in the US, like the early banning of Google Glass users in some San Francisco bars. See Roquet (2016b) for a related analysis of the social ‘seamfulness’ of AR glasses.

- Carmack, John. 2018. Oculus Connect 5 Keynote. September 27, San Jose, California.
- [https://youtu.be/VW6tgBcN\\_fA](https://youtu.be/VW6tgBcN_fA)
- Chambers, Ian. 1990. "A Miniature History of the Walkman." *New Formations* 11: 1-4.
- Chow, Rey. 1990. "Listening Otherwise, Music Miniaturized: A Different Type of Question About Revolution." *Discourse* 13(1): 129-48.
- Crary, Jonathan. 1990. *Techniques of the Observer: On Vision and Modernity in the Nineteenth Century*. Cambridge, MA: MIT Press.
- DeNora, Tia. 2000. *Music in Everyday Life*. Cambridge: Cambridge University Press.
- Dyson, Frances. 1996. "When is the Ear Pierced? The Clashes of Sound, Technology and Cyberculture." In *Immersed in Technology: Art and Virtual Environments*. Mary Anne Moser, ed, 73-101. Cambridge, MA: MIT Press.
- Fink, Robert. 2005. *Repeating Ourselves: American Minimal Music as Cultural Practice*. Berkeley: University of California Press.
- Furness, Thomas. 1988. "Harnessing Virtual Space." *Proceedings of SID International Symposium Digest of Technical Papers*, 4-7. Playa del Rey, CA: Society for Information Display.
- Gay, Paul du, Stuart Hall, Linda Janes, Anders Koed Madsen, Hugh Mackay, and Keith Negus. 2013. *Doing Cultural Studies: The Story of the Sony Walkman*. Second edition. Thousand Oaks, CA: SAGE Publications Ltd.
- Gotō Toshiyuki, Kimura Yōichi, Murai Katsumi, and Yamada Akitoshi. 1976. "Bainōraru juchō no onzō tōgai teii ni okeru zankyōon no yakuwari ni tsuite" [Role of reverberant sound as a factor of 'outer head localization' in binaural headphone listening]. *Chōkaku kenkyūkai shiryō* [Proceedings of the auditory research meeting] 31 (1): 14-17.

- Gotō Toshiyuki, Kimura Yōichi, and Watanabe Kōji. 1976. "Bainōraru rokuonhō" [Binaural recording method]. *National Technical Report* 22 (4): 469–77.
- Grau, Oliver. 2003. *Virtual Art: From Illusion to Immersion*. Leonardo. Cambridge, Mass: MIT Press.
- Groening, Stephen. 2016. "'No One Likes to Be a Captive Audience': Headphones and In-Flight Cinema." *Film History* 28 (3): 114-38.
- Hagood, Mack. 2018. "Here: Active Listening System: Sound Technologies and the Personalization of Listening." In *Appified: Culture in the Age of Apps*. Jeremy Wade Morris and Sarah Murray, eds, 276-85. Ann Arbor: University of Michigan Press.
- . 2019. *Hush: Media and Sonic Self-Control*. Durham, N.C.: Duke University Press.
- Headlam, Bruce. 1999. "Origins: Walkman Sounded Bell for Cyberspace." *New York Times*, July 29, 1999.
- Hosokawa, Shūhei. 2012. "The Walkman Effect." In *The Sound Studies Reader*, edited by Jonathan Sterne, 104–16. New York: Routledge.
- Iwai Seiji. 1955. "Denpa no sekai no miharashidai" [Observation tower on the world of radio waves]. *Denpa kagaku* [Radio science], January 1955. 1.
- Jay, Mark A. 2010. "Holophonic vs. Binaural." <https://www.gearslutz.com/board/all-things-technical/493726-holophonic-vs-binaural.html>.
- Jones, Steven. 1993. "A Sense of Space: VR, Authenticity, and the Aural." *Critical Studies in Mass Communication* 10: 238–52.
- Keightley, Keir. 1996. "'Turn It Down!' She Shrieked: Gender, Domestic Space, and High Fidelity, 1948-59." *Popular Music* 15(2): 149-77.

- Krebs, Stefan. 2017. "The Failure of Binaural Stereo." *ICON: Journal of the International Committee for the History of Technology* 23: 113-43.
- Krier, Beth Ann. 1987. "The Essence of Cocooning." *Los Angeles Times*, August 7, 1987.
- LaBelle, Brandon. 2010. *Acoustic Territories: Sound Culture and Everyday Life*. New York: Bloomsbury.
- Lamarre, Thomas. 2018. *The Anime Ecology: A Genealogy of Television, Animation, and Game Media*. Minneapolis: University Of Minnesota Press.
- Lefebvre, Henri. 2004. *Rhythmanalysis: Space, Time, and Everyday Life*. Trans. Stuart Elden and Gerald Moore. London: Continuum.
- Mannering, Lindsay. 2015. "Now Playing in Your Headphones: Nothing." *New York Times*, December 22.
- Mills, Mara. 2011. "Hearing Aids and the History of Electronic Miniaturization." *IEEE Annals for the History of Computing* 33(2): 24-44.
- Nango Yoshikazu. 2018. *Hitori kūkan no toshiron* [Urban theory of the one person space]. Tokyo: Chikuma shinsho.
- Nishimaki Yasao. 1976. "Watashi no bainōraru rokuon saisei hōshiki no susume to jissai" [My recommendations and the reality of methods of binaural recording playback]." *Rajio gijutsu* [Radio technology], July 1976.
- Okahara Masaru. 1977. "Saikin no bainōraru" [Recent binaural]. *Hōsō gijutsu* [Broadcasting technology], January 1977.
- Roquet, Paul. 2016a. *Ambient Media: Japanese Atmospheres of Self*. Minneapolis: University Of Minnesota Press.

- . 2016b. "From Animation to Augmentation: *Dennō Coil* and the Composited Self." *Animation* 11(3): 228-45.
- Shindenki* [New electronics]. 1955. "Jōhōriron" [Information theory], February 1955.
- Stankieveh, Charles. 2007. "From Stethoscopes to Headphones: An Acoustic Spatialization of Subjectivity." *Leonardo Music Journal* 17 (December): 55–59.
- Sterne, Jonathan. 2003. *The Audible Past: Cultural Origins of Sound Reproduction*. Durham, NC: Duke University Press.
- . 2015. "Space within Space: Artificial Reverb and the Detachable Echo." *Grey Room* 60 (July): 110–31.
- Seibundō shinkōsha. 1962. *Sutereo haifai seisaku dokuhon* [Guidebook to building a stereo hi-fi]. Tokyo: Seibundō shinkōsha.
- Takemura Mitsuhiro. 1987a. "Onzō sōsa, shimiyurakuru no rinkaiten: Seitai media to 'horofonikusu'" [Sound image scanning, simulacra at a critical point: Biomedica and 'holophonics']. *Yuriika* [Eureka], June 1987.
- . 1987b. "'Genjitsu' o yurugaseru oto: Onzō shimiyurēshon 'horofonikusu'" [The sound that shook 'reality': Sound image simulation 'holophonics']. *Kōkoku* [Advertising], September 1987.
- Théberge, Paul, Kyle Devine, and Tom Everett, eds. 2015. *Living Stereo: Histories and Cultures of Multichannel Sound*. New York: Bloomsbury Academic.
- Thibaud, Jean-Paul. 2003. "The Sonic Composition of the City." In *The Auditory Culture Reader*, edited by Michael Bull and Les Back, 329–42. New York: Berg.
- Thompson, Emily Ann. 2002. *The Soundscape of Modernity: Architectural Acoustics and the Culture of Listening in America, 1900-1933*. Cambridge, Mass: MIT Press.

- Tsuzuki Kyōichi. 1993. *Tokyo Style*. Tokyo: Chikuma shobō.
- Tuhus-Dubrow, Rebecca. 2017. *Personal Stereo*. New York: Bloomsbury Academic.
- Wakabayashi Shunsuke. 1961. “Iyāfōn!?: Sutereo saisei to bainōraru saisei no chigai, sono 1” [Earphones!?: The difference between stereo playback and binaural playback, part 1]. *Rekōdo geijutsu* [Record art], April 1961.
- Weber, Heike. 2010. “Head Cocoons: A Sensori-Social History of Earphone Use in West Germany, 1950–2010.” *The Senses and Society* 5 (3): 339–63.
- Weheliye, Alexander G. 2005. *Phonographies: Grooves in Sonic Afro-Modernity*. Durham, N.C.: Duke University Press.
- Williamson, Judith. 1986. *Consuming Passions: The Dynamics of Popular Culture*. London: Marion Boyars.
- Wurtzler, Steve. 2008. *Electric Sounds: Technological Change and the Rise of Corporate Mass Media*. New York: Columbia University Press.
- Yamada Akitoshi and Kikuchi Yoshinobu. 1976. “Tekunikusu ambiensuhon no tokuchō” [Characteristics of the Technics Ambience Phone]. *Rajio gijutsu* [Radio technology], July 1976.
- Yamamoto Seishū. 1975. “Heddohon o rikai suru kōza” [Course on understanding headphones]. *Rajio gijutsu* [Radio technology], February 1975.
- Yasar, Kerim. 2018. *Electrified Voices: How the Telephone, Phonograph, and Radio Shaped Modern Japan, 1868–1945*. New York: Columbia University Press.
- Yoshimi Shunya. 2012. “Koe” no shihonshugi: denwa, rajio, chikuonki no shakaishi [Capitalism of the “voice”: social history of the telephone, radio, and phonograph]. Tokyo: Kawade shobō shinsha.





