

HISTORY OF 3D SOUND

BRAXTON BOREN

American University

Introduction

The history of 3D sound is complicated by the fact that, despite how much the concept may appear to be a late-20th century technological buzzword, it is not at all new. Indeed, just as Jens Blauert famously reminded us that “there is no non-spatial hearing” (Blauert, 1997), so too due to the nature of our world all sound is inherently three-dimensional (Begault, 2000). For the majority of human history the listener – the hunter in the field, a singing congregant within a cavernous stone church, or the audience member at a live performance – perceived sound concomitantly with its spatial setting.

In this sense it is the late-20th century view that is out of step with historical sound perception.

The advent of audio recording in the 19th century led to the development of zero-dimensional (mono) sound, and later one-dimensional (stereo), and two-dimensional (quad and other surround formats) reproduction techniques. Due to the greater sensitivity of the human auditory system along the horizontal plane, early technology understandably focused on this domain. Our capability to mechanically synthesize full 3D auditory environments is relatively recent, compared to our long history of shaping sound content and performance spaces.

The effect of physical space is not limited to the perceived physical locations of sounds – different spaces can also affect music in the time domain (e.g. late reflection paths) or frequency

domain (by filtering out high-frequency content). Often a listener's experience of a space's effect on sound – such as singing in the shower or listening to a choir in a reverberant stone church – is describing primarily non-localized qualities, which could be captured more or less in a monaural recording. Though space has always been an integral part of live performance, it has rarely served as more than an additional ornamentation on this spectral/temporal palette for most composers (with some notable exceptions), though technological advances are leading to progress on this front.

The other area focusing on 3D sound currently is the field of virtual auditory spaces (VAS), which may be used for either virtual reality simulations or augmented reality integrations of 3D sound into the listener's existing auditory environment. In contrast to the musical front, these applications' goals fundamentally require convincing spatial immersion, elevating the importance of space above that of frequency or time in many cases. In a sense, the rapid developments in these fields seek to re-establish that connection between sound and space, which was to some extent severed by early audio recording and reproduction. It follows then, that to look forward to the future of 3D sound, we should first look backward to the various uses and experiences of sound and space throughout history.

Prehistory

Because sound is inherently transient, most sounds of the past are lost to present day observers. For the prehistoric period – that is, the time before written language or history – we lack even subjective descriptions of sounds and thus must rely on the tools of archaeology to reconstruct the acoustic world experienced by early humans. Hope Bagenal famously stated that the acoustics of all auditoria originated from either the open air or the cave (Bagenal, 1951). For the

case of our earliest ancestors this was literally true as they spent most of their time hunting and gathering in outdoor environments that approximated free field listening conditions. Their 3D sound localization was honed in this largely nonreverberant context, allowing them to evade predators and find their own prey.

However, when these early hunter-gatherers stepped inside the diffuse sound field of a cave, they would have found an acoustic environment wholly different from that outside. The reflective walls and enclosed spaces would have generated audible echoes, modal resonances, and reverberation – causing listeners to be surrounded by thousands of copies of their own voices, an immersive, amplified, and somewhat mystical experience. Even today when we are more used to moderate amounts of reverb in popular music, entering a large stone church with a long reverberation time for the first time yields a sense of 3D envelopment unrivaled by the best surround sound system. For our prehistoric ancestors who had no such experience or knowledge, such a space would have sounded otherworldly – a complete removal from the world they knew outside.

How did this experience shape the early humans' use of caves? Analysis of Paleolithic caves in Britain, Ireland, and France suggests many connections and correlations between the locations of rock drawings and strong acoustic resonances, particularly those in the range of the male singing voice (Jahn *et al.*, 1996; Reznikoff, 2008). The discovery of twenty shell trumpets at Chavín de Huántar in Peru suggests that this space was used for musical performance during ritual ceremonies. Because Chavín consists primarily of stone-cut galleries (constructed around 600 B.C.), these smaller interior spaces possess lower reverberation times but still provide acoustic immersion because of widely distributed reflection patterns and noncoherent energy density

(Abel et al., 2008). This unique acoustic environment, which does not occur in either the free field or most natural caves, can be seen as a precursor to the modern concert hall, which minimizes the correlation of the pressure signals at both ears, resulting in an enveloping sound field, while still maintaining sufficient acoustic clarity to understand the content of the music being played (Kendall, 1995b).

Ancient History

With the transition from nomadic hunter-gatherers to settled agricultural societies, architectural spaces also transitioned from natural or roughly-hewn stone caves to more advanced ceremonial spaces. The open-air condition evolved into the Greek amphitheaters, which had roots in the Minoan theaters from as early as 20,000 B.C. but found their apotheosis in the Classical amphitheaters around the 5th century B.C. (Chourmouziadou & Kang, 2008). These spaces set the standard for theater design in Western culture, maintaining excellent speech intelligibility through strong early reflections while avoiding confusion stemming from reverberation. Recent analysis has also shown that diffraction effects from the corrugated seating structure in Greek amphitheaters preferentially amplify the frequencies associated with speech intelligibility while attenuating low frequency noise (Declercq & Dekeyser, 2007). Vitruvius, a Roman architect whose writings are the best preservation of Roman and Greek architectural knowledge, mentions that in some theaters bronze vessels were placed beneath the seats whose resonances amplified the actors' speaking voices (Vitruvius, 1914). Though it is doubtful Vitruvius actually saw these vessels in action (Rindel, 2011b), the approach is a striking early example of distributing sound sources throughout a performance space.

As the Classical theaters gave way to Hellenistic and later Roman theaters, changes were made that added reverberation and decreased intelligibility (Chourmouziadou & Kang, 2008; Farnetani, Prodi, & Pompoli, 2008). This in effect made these outdoor theaters somewhat of a bridge between Classical amphitheaters and *odea*, the interior halls built throughout Greece specifically for the performance of music (Navarro, Sendra, & Muñoz, 2009). These featured higher reverberation times and lower musical clarity, but greater sound reinforcement for relatively weak instruments such as the Greek lyre (Rindel, 2011a). The Greek *odeon* was a singular example of architecture designed around music, as this pattern was not generally followed throughout the rest of the ancient world: elsewhere music had to adapt itself to performances in theatres and other public spaces that were constructed based on non-musical criteria, such as optimal speech intelligibility, and sometimes non-acoustic criteria, such as maximizing seating or using the cheapest materials possible.

Because architectural acoustics as a discipline was largely developed in the West, so too the majority of historical acoustical analysis has focused on Western civilization from Greek antiquity onward. However, enclosed immersive temples are found throughout the world, and recent analysis has begun to examine acoustic phenomena particular to non-Western cultural and religious traditions (Prasad & Rajavel, 2013; Soeta, Shimokura, Kim, Tomohiro, & Ken, 2013). Meanwhile, in the West the Christian church would serve as the primary site of musical performance for over a thousand years (Navarro et al., 2009).

Space and Polyphony

To understand the effect of space on musical development in the first millennium A.D., it is first helpful to understand the basic context surrounding the rise of the Christian church, which so

profoundly shaped music composition during this period. Originally a splinter sect of Judaism, Christian worship had its roots in the Jewish synagogue, which focused on readings and exhortations, which were generally spoken or chanted as a monotone (Burkholder, Grout, & Palisca, 2006). Early Christian worship retained this spoken liturgy, which was appropriate for their physical setting: since Christians refused to worship the Roman emperor, they were not allowed recognition by the empire and thus met in small groups within house churches whose dry acoustics matched their spoken liturgy. However, after the Emperor Constantine issued the Edict of Milan in 313, confiscated Christian property was returned, and Roman architects began building large stone basilicas for Christian worship with long reverberation times. Around this same point the Christian liturgy became more focused on a sung liturgy, which slowed the rate of spectral variation over time in the highly time-dispersive communication channel from the priest to the congregation (Lubman & Kiser, 2001). With this change we see the first large swing of the history of Western music on Bagenal's continuum: while the house churches had resembled the clarity of the "open air," the church now found itself in the "cave," which shifted worship away from the semantic content of the spoken word toward the aesthetic experience of being surrounded by a line of chanted music and its many reflections.

As the Christian church grew in size and influence, many varieties of sung chant took root, but the best known repertory, codified in the early eighth century, was known as Gregorian chant. Though attributed to Pope Gregory I (r. 590-604) it is more likely that the standardization of chant took place under Pope Gregory II (r. 715-31) (Burkholder et al., 2006). The proto-Romanesque churches being built during this period were made of reflective stone materials and possessed large volumes that ensured that any monophonic sung line would be heard along with the reflection of one or more previous tones. Navarro argues that over time this "ensured a

certain education and familiarization with polyphonic sound. Indeed, the persistence of the sound of the different notes of the melody [led] to melodic developments, structured in such ways that the simultaneity of notes [was] resolved over certain harmonies” (Navarro et al., 2009). Lubman and Kiser (Lubman & Kiser, 2001) go even further and argue that the piling up of multiple notes at once in reverberant churches was a catalyst for the development of polyphonic music in the West.

The simplest polyphony – that of a single ‘drone’ under the melody – had existed in both the East and the West since antiquity. However, the earliest music with multiple moving parts is known generally as organum. The simplest and earliest form of organum, in which a parallel tone was sung along with the main melody of the chant, was apparently already a standard practice by the ninth century (Burkholder et al., 2006). Indeed, the term organum does not refer to polyphony *per se*, but rather to voices which co-existed in a way consistent with natural law (Fuller, 1981). Indeed, modern listeners might instead use the word ‘organic’ to describe this relationship, which also suggests an early connection between organum and reverberation, which blends direct and indirect sound into a single perceptual object. Without direct documentary evidence, the hypothesis of modern polyphony’s origins in reflected sound cannot be proved explicitly, yet it is still a compelling candidate for the first experimental environment for the comparison of multiple tones moving at once. If true, this would indicate that the immersive component of sound has affected not only our spatial arrangements of music but also the very fabric of how music theory itself has developed.

Even after the development of polyphony, church design continued to strongly influence music composition: the Gothic churches, which followed the Romanesque period often possessed a

single resonance that was known as the ‘note’ of the church. Since institutional Christianity at this point generally opposed instrumental accompaniment during the Mass (Burkholder et al., 2006), the church’s natural ‘note’ served as their reference tone and thus the church itself was a sort of natural accompaniment to the *a cappella* vocal music by reinforcing the choir within a single key (Bagenal, 1951). A more detailed history of the influence of architectural space on Western music composition is given in (Forsyth, 1985).

Spatial Separation in the Renaissance

The word ‘renaissance’ denotes *rebirth*, and was meant to signify the rediscovery of classical Greek and Roman culture after the supposed backwardness of the Middle Ages. At least in the case of music, this narrative is not accurate: as we have already seen, the Medieval organum style expanded upon the musical traditions of the classical world, and during the Renaissance advances in polyphony would surpass both the Medieval and classical styles in scope and complexity. Indeed the complex polyphony of the Venetian Renaissance seems at first glance to be rather ill-suited to the large reverberant churches Palladio and other architects designed in Venice during this period (Howard & Moretti, 2010). However, computational simulations show that on the festive occasions for which such music was composed, large crowds and wall tapestries could have reduced the reverberation by as much as a factor of one half, drastically increasing the clarity of the performance (Boren & Longair, 2011; Boren, Longair, & Orłowski, 2013).

But despite the advances in polyphonic style during this period, perhaps more significant to our modern ears was the practice of *cori spezzati*, the composition of music for multiple choirs that were separated in space. This practice originated in the late 15th century in northern Italy, spread

throughout the region in the early 1500s, and came to its apex in Venice under Adrian Willaert, Andrea Gabrieli, and Claudio Monteverdi (Arnold, 1959; Howard & Moretti, 2010). Willaert was the first major composer to adopt this style, and the considerations he made in his compositions for this ensemble resemble those made by an engineer mixing a stereo recording: he made sure to include a wide spectral range in both choirs in case a listener was too near a single choir, and he implemented one of the earliest documented uses of doubled bass lines to keep his ensembles together. Good examples of this polychoral style include Willaert's *Vespers* of 1550 and Gabrieli's 3-choir mass for the visit of the ambassadors of Japan in 1585 (Zvonar, 2006; Arnold, 1959). Again we see that independent of localization effects, spatial factors significantly affected the tonal development of Western music even at this early juncture.

But besides these tonal effects, it seems clear that the spatialization of separated choirs was an integral part of the aesthetic of *coro spezzato* music. Simple call-and-response antiphony, both recited and sung, dates back to antiquity (Slotki, 1936), but the more complex composition of music for spatially separated ensembles was fully realized for the first time in the Venetian Renaissance. Though we have evidence that on certain occasions the choirs performed in a single location (Fenlon, 2006), Gioseffo's Zarlino's comments on the genre suggest that physical separation of the choirs "is a structural requisite for this particular genre of composition...and not merely one of various possibilities" (Moretti, 2004). Because the ruler of Venice, Doge Andrea Gritti, became too fat to reach his old elevated throne in the Basilica San Marco, in 1530 he moved his seat into the chancel, where previously the high priest had resided (Howard & Moretti, 2010). After this move, Jacopo Sansovino, the chief architect of the church, constructed two *pergoli*, or raised singing galleries, on either side of the doge's new throne because former galleries lower down had become obstructed by wooden seating in the chancel. Moretti argues

that these galleries were used to give a stereo effect for the performance of split-choir music to the doge's position (Moretti, 2004). On-site acoustic measurements showed that these positions produced a near-perfect blend of clarity and reverberation at the doge's position, while other congregants in the nave received a much muddier sound (Howard & Moretti, 2010). Further analysis showed that this effect resulted from Sansovino's galleries maintaining a direct line-of-sight to the doge's position, without which the galleries would instead produce the same unclear acoustics heard by the parishioners (Boren et al., 2013).

Outside of Italy, one notable polychoral work from this period is Thomas Tallis's *Spem in alium*, a 40-piece motet composed in England around 1570 for eight separate choirs of five voices each. Tallis may have been inspired to compose this piece by a similar 40-voice motet by Alessandro Striggio, an Italian composer who visited London in 1567. Many details of the original performance of Tallis's piece are unknown, but it is thought to have been performed for the first time in Arundel House, London, around 1571 (Stevens, 1982). Though we have no indications of the spatial arrangement of the choirs on this occasion, such a large array of choirs inevitably would have created some spatial separation. At least by 1958 *Spem in alium* had been staged with the eight choirs arranged in a circle, and the audience enclosed in the center (Brant, 1978). Tallis's motet, however, is most significant to the history of spatial sound because it served as the basis for a 2001 sound installation, *40 Part Motet*, by Janet Cardiff. This exhibit featured a 40-channel close-miked recording of each of the voices of Tallis's motet, played on 40 raised loudspeakers arranged in a circle (MacDonald, 2009). This installation, which debuted when most consumers had access to only 5-channel surround sound, was one of the earliest exposures of many outside the professional audio community to the possibilities of simulating the full sound field of a real performance distributed in space.

Spatial Innovations in Acoustic Music

Baroque Period

At the dawn of the Baroque period (1600-1750), modern-day Germany was in a state of transition due to the success of the Lutheran Reformation, begun by Martin Luther in 1517. This catalyzed a conflict between the Catholic Church and the Lutherans, who stressed, among other things, the importance of preaching and singing in the vernacular German language, since many congregants could not understand Latin. This led the Reformers to stress the importance of clarity and speech intelligibility in their churches: as formerly-Catholic churches were taken over by Lutherans, the spaces were altered to improve the clarity of the spoken word. After the Lutherans took over the Thomaskirche in Leipzig in 1539, they made a variety of alterations to the space that greatly reduced the reverberation (Lubman & Kiser, 2001) and made it more like “a religious opera house” (Bagenal, 1930). This church, of course, is most famous as the compositional home of Johann Sebastian Bach (1685-1750) from 1723 until the end of his life. This extreme swing away from the “cave” and back toward the “open air” thus led to “the acoustic conditions that made possible the seventeenth century development of cantata and Passion” (Bagenal, 1930). While the more reverberant churches had been appropriate for chant music at a slow, steady tempo, the drier acoustics introduced by the Reformation allowed Bach to make use of dramatic shifts in tempo with works such as the famous St. Matthew Passion (Bagenal, 1951). Thus theological considerations shaped architectural development, which in turn shaped musical development through the singular musical career of Bach, one of the most influential composers in Western history.

Meanwhile, in the southern Catholic regions of Europe, the situation was quite different but no less spatially interesting: it is interesting to note that St. Peter's Basilica in Rome, whose immense expense necessitated the selling of indulgences that spurred Luther to write his 95 theses, later provided the space necessary for the immense scale of the "Colossal Baroque" style, which featured performances by twelve separate choirs, each of which had a separate organ accompanying it (Dixon, 1979). The best known example of this opulent genre is the late-17th century *Missa Salisburgensis* by Heinrich Biber, which featured five separated choirs, each with different accompanying instruments, as well as two separated ensembles of brass and timpani positioned in the galleries of the Salzburg Cathedral (Hintermaier, 1975; Holman, 1994). Whereas in Leipzig the drier acoustics allowed Bach to explore music in tonal and temporal dimensions, in the less clear churches of southern Europe, physical separation helped listeners perceive the different ensembles despite the spaces' longer reverberation.

Classical Period

It is tempting to hypothesize the dramatic spatial effects that might have been used by the Salzburg Cathedral's most famous composer, Wolfgang Mozart (1756-1791), when he wrote for the space a century after Biber. However, perhaps in reaction to the dramatic spatial effects employed during the Baroque period, the Prince-Archbishop of Salzburg issued a decree in 1729 that forbade the wide separation of ensembles and confined all music to the main choir in front (Rosenthal & Mendel, 1941). Visitors to the Salzburg Cathedral often remark at the site of the church's five organs: four near the chancel in the front (though not the same the organs from Mozart's time), and the larger main organ at the rear. Yet based on Mozart's father's account of performance practice, it does not seem likely that multiple organs were used simultaneously

(Harmon, 1970). Though Mozart was unable to experiment within the interior setting, he did employ spatial separation in his secular music: in the score for his opera *Don Giovanni* (1787), Mozart wrote parts for three separate orchestras: one in the pit, one onstage, and one backstage. Each ensemble plays highly differentiated material in separate meters, requiring very precise temporal coordination (Brant, 1978). Mozart also employed a larger distance separation in his *Serenade for Four Orchestras*, K. 286, which likely composed for an outdoor gathering in Salzburg (Huscher, 2006). This work employs an echo effect between the separate orchestras that can be somewhat confusing in a reverberant interior hall, but is well-suited to a free-field outdoor performance, since each orchestra provides ‘reflections’ for the others’ initial motives.

Echo effects were a popular spatial effect during the Classical period – Mozart’s friend Joseph Haydn (1732-1809) also employed non-spatially separated echoes in the second movement of his Symphony No. 38, sometimes called the “Echo” symphony for this reason. In this case the echo originates from violins played at normal strength, followed by muted violins. Haydn elaborates this concept more fully in “Das Echo” (Hob. II:39), a string sextet for two trios spatially separated, traditionally in different rooms (Begault, 2000). In this piece Haydn changes the length of time before the echo is heard, beginning with a whole measure delay, then a half note delay, quarter note delay, and eventually the echo is shortened to only an eighth note. This has the aesthetic effect of changing not the spacing of the ensemble but the size of the virtual room Haydn is simulating – perhaps the earliest known example of altering a performance space dynamically as a musical parameter. Echo effects can still be heard today, either real or simulated, at many funerals: the playing of Taps issues from a single bugle, and then is heard from another bugle, farther away. In areas where horn players are scarce, many a bugler has had

to perform a simulated echo by either turning around and playing away from the funeral site or quickly running to the other side of a hill and playing again.

Romantic Period

While the Classical period used echo effects as a sedate gesture towards an abstract acoustic space, the Romantic trend toward programmatic music that told a distinct story also led to more radical uses of spatialization. Perhaps no composer would make better use of spatial storytelling than that master of Romantic program music, Hector Berlioz (1803-1869). Earlier, François Joseph Gossec had surprised and alarmed the audience for his *Grande messe des morts* (1760) by using a separate brass ensemble hidden in upper reaches of the church, which suddenly gave the call of the last judgment. This is thought to have inspired Berlioz to go further and use four antiphonal brass choirs, each placed in one of the cardinal directions, to sound the call to judgment in his *Requiem* in 1837 (Broderick, 2012). Instead of a surprising call from afar, the audience is surrounded by the brass ensembles' sharp attacks that give illusions of surrounding the audience through the time differences between ensembles as well as the inevitable reflections of each ensemble through the performance space. In fact, we know that Berlioz was aware of this spatial effect as well, for he wrote only two years earlier

Many fail to recognize that the very building in which music is made is itself a musical instrument, that it is to the performers what the sound board is to the strings of the violin, viola, cello, bass, harp, and piano stretched above. (Bloom, 1998)

Thus it seems that Berlioz conceived the *Requiem* not only as a set of sound sources in space, but rather as a single 3D immersive environment whose spatial characteristics could be controlled to some extent through careful orchestration and ensemble placement.

Later in his *Symphonie Fantastique*, Berlioz uses offstage instruments and a very specific narrative program to prime his audience to hear music “here” and “there” (Begault, 2000). His offstage oboe in the third movement echoes back the tune of the English horn, representing two shepherds piping to each other across a valley. In addition to the spectral low-pass effect of being behind the stage curtain, the oboe also makes a slight change to the initial theme, a clear demarcation between the shepherd nearby and his friend far-off. At the end of the movement, the horn repeats its call but is not answered. This primes the audience to listen for something far away, and instead of the friendly oboe they are instead greeted with the ominous terror of the famous March to the Scaffold in movement 4, as the executioner’s procession begins to approach the protagonist (Ritchey, 2010). Berlioz’s offstage invocation of a far-off sound source would also be adopted late Romantic composers such as Giuseppe Verdi (1813-1901), who used an offstage ensemble in his own *Requiem* (1874), and Gustav Mahler (1860-1911), who used offstage brass in 1895 at the premiere of his “Resurrection” Symphony No. 2 (Zvonar, 2006).

20th Century Acoustic Music

Though many of the contemporary advances in spatialization occurred through technological advances, there were also some 20th-century composers who continued in the purely acoustic spatial tradition that dated back to Willaert and the Renaissance. Thus it will be more helpful to cover these composers first, and then discuss electroacoustic spatialization after the history of spatial technology from the 19th century onward. The spatial storytelling of the Romantic Era was continued within the American Experimental movement, as typified by Charles Ives (1874-1954).

Charles Ives grew up learning from his father George Ives, who was a bandmaster during the American Civil War and had much experience with another archetypal form of moving sound

sources: the marching band. Indeed, George at one point conducted an experiment in which he led two separate bands, marching in opposite directions through the town square (Zvonar, 2006). His son Charles, known for his juxtapositions of contrasting musical material, would add spatial separation to his compositional toolkit in *The Unanswered Question* (1908). In this piece Ives places a trumpet and woodwinds onstage, which respectively pose “The Perennial Question of Existence” and various answers. Meanwhile, Ives places a separate string quartet offstage, which represents “The Silences of the Druids—who Know, See and Hear Nothing,” and more broadly can be seen to be “representatives of the unfathomable cosmos beyond” (McDonald, 2004). Thus Ives uses spatial separation as a stand-in for the cosmic separation between Us – the artists, the thinkers, the ones asking and answering questions – and It – the cosmos, which will go on in silence long after we have ceased our questioning. A later performance of the piece by Henry Brant also separated the Questioner and the Answerers, perhaps indicating some metaphysical distance between those personalities as well (Brant, 1978).

Brant (1913-2008) was strongly influenced by Ives’s use of space, particularly in *The Unanswered Question*. Whereas the previous examples have largely been major composers for whom space was a minor effect, Brant is in many ways the opposite: while not a dominant figure in 20th century music, Brant’s body of work embraced and explored spatial music more than any acoustic composer before him. Beginning with *Antiphony I* (1953), Brant would compose spatially organized music for separated ensembles for the next fifty years (Harley, 1997). Though he admitted that electric reproduction could add flexibility to spatial performance, Brant disliked loudspeakers because their directivity was markedly different from the live instrumentalists or vocalists (Brant, 1978). Where Ives indicated a general separation between ensembles, Brant rigorously specified positions for different ensembles within a performance space (Harley, 1997).

Though his compositional contribution is significant, Brant was also active as a theorist of spatial music theory. He believed that physical space provided freedom to the composer by allowing an ensemble to be more compressed in tonal space: whereas in a single ensemble a unison between different instruments led to confusion, when those instruments are spread out a shared tone may be a benefit rather than a hindrance (Brant, 1978). Brant possessed no formal scientific education, but he was aware of many critical psychoacoustic concepts through a lifetime of musical experiments. In particular, he, like Ives before him, alludes to the “Cocktail Party Effect,” whereby a listener may more easily shift attention between highly contrasting material when the sound sources are spatially separated (Harley, 1997). In addition, Brant anticipates later research on localization blur, when he suggests that loudspeakers’ directivity is less of a problem when placed on the ceiling, and also the spectral variation of the head-related transfer function (HRTF), when he states that high and low pitches are “enhanced” by high and low elevations, respectively (Brant, 1978). Though his music is called ‘spatial,’ Brant’s writings do not indicate that space is the primary organizational dimension for his work. Rather, he consistently saw space from a utilitarian perspective, whereby traditional constraints of tonal composition could be alleviated through a more expansive exploration of physical space.

3D Sound Technology

By the time we reach the 19th century, the history of 3D sound begins to be tied more and more closely to the development of modern science and technology. Though research into acoustics dates back to antiquity (Lindsay, 1966), a rigorous theory of sound localization in 2D would not be put forth until the late 19th century and gradually refined for a century afterward to account for 3D localization (Strutt, 1875; Blauert, 1997; Kendall, 1995a). However, as is often the case, the

technology necessary to achieve spatial auditory effects often plunged forward far before the science behind the process was well understood.

Binaural or Stereo?

It may be argued that 3D audio technology begins and ends with binaural, though as we shall see, we must take some care with how we define the term. The word ‘binaural’ refers, at the most basic level, to hearing with two ears, but it later came to include all the spatial cues from the ears, head, and body of a listener. This odd trajectory stems from the fact that binaural audio is perhaps the easiest spatial effect to capture, but the hardest to realize in post-production. Only four years after Alexander Graham Bell’s invention of the telephone, he did some early experiments using two telephones receivers and transmitters (Davis, 2003). The next year a French engineer named Clément Ader devised a system for the spatial transmission of the Paris Opera over a mile away to a listening booth at the International Exposition of Electricity (Hospitalier, 1881; MacGowan, 1957; Torick, 1998; Paul, 2009). This invention, dubbed the ‘Theatrophone’ used an array of pairs of transmitters across the stage that were then routed to pairs of telephone receivers at the Exposition’s listening booth. Attendees held both receivers to their ears and were able to perceive the spatial position of sound sources through the interaural differences transmitted over the lines. Though the system suffered from insufficient amplification and vibration damping, the service proved to be successful enough to merit a home-subscription service. Among the most well-known Theatrophone subscribers were Marcel Proust and Great Britain’s Queen Victoria. Though popular among the well-to-do in the early 20th century, the advent of cheaper monophonic wireless broadcasting opened up a narrower sound to a wider segment of the population, and Theatrophone broadcasts ceased in 1932. It would be another 30

years before stereo broadcasting would bring back this basic spatial feature that had been discovered so early on (Collins, 2008).

In discussing Ader's significance, it is useful to consider what his technology actually represented: because it transmitted many points along the wavefronts emitting from the Opera's stage, it could be thought of as an early form of wave field synthesis. However, because it conveyed interaural differences to be conveyed to both ears of the listener, some have classified this as the earliest example of 'binaural' sound (Sunier, 1986), or as it was called at the time, 'binauricular,' perhaps too much of a tongue-twister to attain popular acceptance (Collins, 2008). It is important to note that at the time, binaural was principally used to mean hearing with two ears, rather than recording with a real or synthetic human head. The modern distinction between binaural and stereo was not even suggested until the 1930s, and widespread usage of these separate definitions would not follow until the 1970s (Paul, 2009). Thus it is probably safest to say that Ader's achievement was the earliest reproduction of binaural sound under the 19th century understanding of the term. Since the transmissions did not make use of a dummy head to obtain level differences, time differences, or spectral cues corresponding to the actual filtering effects of the head, a present-day understanding might instead classify the Theatrophone as a very effective form of 2-channel stereo, distributed over several listening points.

Despite this distinction, advances in stereophony and proper binaural sound were fast-coming. By World War I (1914-1918), two-ear listening devices were being used to track enemy planes (Sunier, 1986) and also to track submarines using inputs from dual hydrophones (Lamson, 1930).

Some have claimed that artificial heads were used as early as 1886 at Bell Laboratories, but this seems doubtful for several reasons and has not been confirmed (Paul, 2009). Indeed, the earliest definite cases of binaural transmissions using some form of primitive artificial head were both patented in 1927, one by Harvey Fletcher and Leon Sivian, and another system for recording and reproduction by W. Bartlett Jones (Fletcher & Sivian, 1927; Paul, 2009). These both used very basic spheroid objects as a dummy-head, but Fletcher's research at Bell Labs would later develop a more sophisticated binaural recording device in 1931 using a tailor's manikin nicknamed 'Oscar.' The 1.4-inch microphones placed in Oscar were too large to fit into the ear canal, so the microphones were mounted instead on the manikin's cheekbones directly in front of the ears. Listeners to the transmissions from Oscar from the Philadelphia Academy of Music were astounded at the degree of localization that could be achieved – Fletcher stated that “the mechanism by which this location is accomplished is not altogether understood, but the interaction of the two ears seems to have much to do with it, for stopping up one ear destroys the ability almost completely.” Despite not fully understanding the mechanisms of spatial hearing, there was widespread agreement that binaural listening was more pleasant: over a third of listeners in Fletcher's experiments preferred binaural to monaural even when the binaural content was low-pass filtered with a cutoff frequency of 2.8 kHz (Fletcher, 1933). Oscar was later used at an exhibition at the Chicago World's Fair, where listeners were amazed at being able to hear moving sources when there were none around them (Paul, 2009). Despite Fletcher's bold statement that “there is no longer any limitation, except expense, to the acoustic fidelity which electrical transmission systems can achieve” (Fletcher, 1933), there were in fact many front-back confusions and distance errors, as would be expected with a non-individualized static binaural transmission, especially given the limitation effects of the pinnae at Oscar's microphones. Later

famous binaural dummy heads such as KEMAR and the Neumann KU-100 would be developed, along with many others, but despite many advances binaural technology remained a niche area of audio for most of the 20th century. A detailed history of binaural recording devices is given by Stephan Paul (2009).

Loudspeakers: From Stereo to Multichannel

While the recording methods of these early ‘binaural’ methods varied, they had in common an early headphone-based reproduction that allowed most of the coarse interaural differences to be conveyed directly to the ear canals of listeners. However, loudspeaker technology was also rapidly developing during this time, and many spatial experiments would be carried out during this era. As early as 1911 Edward Amet filed a patent on a device to pan a mono record, synchronized with a film projector, around a series of loudspeakers such that the sound of an actor’s voice would follow his position on the screen (Amet, 1911). As Davis notes, “this was a remarkably far-sighted invention, as it would be another dozen years before even mono-synchronized sound was commercially employed in the cinema” (Davis, 2003). Thomas Edison’s phonograph was used for strictly monophonic reproduction, but an audience in 1916 perceived it as very realistic, perhaps due in part to the reverberant acoustics of Carnegie Hall where the demonstration took place (Davis, 2003).

Efforts to store and reproduce stereo sound abounded, partly due to the success of the various binaural experiments mentioned above. A radio engineer in Connecticut named Franklin Doolittle filed patents for 2-channel recording (1921) and broadcasting (1924), and the radio station he owned, WPAJ, began broadcasting sound captured with two microphones, emitting on two separate radio frequencies (Paul, 2009). In 1931, British engineer Alan Blumlein (1903-

1942) filed a patent that is widely considered to mark the birth of stereo, as we understand it today (Blumlein, 1931). As we have already seen, Blumlein was not the first to record or broadcast two audio channels at once, though his patent covered both of these things. The impact of Blumlein's patent should instead be seen in the comprehensiveness with which he envisioned the transformations that stereo sound could bring to the world of audio. Especially innovative were his creation of the 90-degree XY stereo microphone technique (the so-called "Blumlein pair"), his system of amplitude panning between two output channels, and a special disk-cutting technique for recording two channels of audio into either side of a single groove of a record.

Blumlein's work was so far ahead of its time that when he died in an airplane crash at the age of 38, he and his work were still largely unknown. His stereo disk-cutting technique, relatively unknown and ahead of its time, would later be re-invented twice more, in separate patents by Bell Labs and Westrex Corporation respectively, before becoming commercially viable (Davis, 2003).

As even a monophonic wireless radio was a big investment for the average household at this time, much of the commercial research into multichannel audio reproduction came from motion picture companies who could afford to invest in cutting-edge technology to draw listeners to a media experience that surpassed anything they could hear at home. However, many early attempts made use of expensive prototypes and were often abandoned afterwards. In New York City, a rehearsal screening of *Fox Movietone Follies of 1929* made use of the same concept that Amet patented 18 years before: a monitoring device was used to pan the monophonic movie sound track back and forth between the left and right loudspeakers, but after this, Fox gave up on the idea (MacGowan, 1957). Conductor Leopold Stokowski, who had been involved with

Harvey Fletcher's earlier experiments using the binaural dummy Oscar, also worked with Fletcher in 1933 to produce a three-channel transmission of the Philadelphia Orchestra, reproduced for an audience in Washington, D.C. (Torick, 1998). This three-microphone-to-three-loudspeaker recreation was a more perfect antecedent to wave field synthesis even than was Clement's Theatrophone. But Stokowski's experiment would also pave the way for the first use of surround sound in film through the 1940 Walt Disney film *Fantasia*, for which Stokowski was the conductor (MacGowan, 1957). For this occasion a new audio system was designed called Fantasound, which used three tracks of audio similar to that transmitted during the 1933 experiment. However, the system also used a separate optical control track to pan the three audio tracks to any of ten groups of loudspeakers: nine surrounding the audience horizontally and one on the ceiling (Torick, 1998). The rear and elevated speakers were not used excessively, but during the film's finale, Franz Schubert's *Ave Maria*, they were used to give the sensation of a chorus progressing from the rear to the front of the audience, yielding perhaps the most complete surround immersion yet achieved in a commercial audio system (Malham & Anthony, 1995). Yet again, the system was largely abandoned after this technical achievement, and the playback equipment was unfortunately lost at sea thereafter (Davis, 2003). Stokowski remained an avid believer in audio exploration for the rest of his life, proclaiming to the Audio Engineering Society in 1964 that audio technology represented "the greatest mass presentation of musical experience so far known to man" (Torick, 1998). Though he had no technical training, Stokowski's reputation and influence greatly aided Fletcher and other audio scientists and engineers during the early development of 3D sound technology.

As stereo began to reach commercial viability in the late 1950s, work was already progressing on more ambitious multichannel audio formats. By the early 1960s commercial efforts had been

made to market a system that extracted out-of-phase signal components from a standard stereo recording to a separate pair of rear loudspeakers. By 1968 the earliest ‘quad’ system was proposed by Peter Scheiber, who developed a system to compress four analog channels into just two for storage purposes, while reconstructing the original four channels under certain constraints of channel separation and phase artifacts (Torick, 1998; Davis, 2003). A variety of quad matrixing formats followed, possibly motivated by “commercial one-upmanship, [which] arguably result[ed] in products and systems being rushed into the marketplace prematurely” (Davis, 2003). At any rate, despite the aggressive marketing of quad formats throughout the 1970s, the format failed to attain commercial success, resulting in a widespread loss of confidence about the future of 3D sound technology.

However, out of the ashes of the commercial failure of these early quadrophonic systems rose several technologies that would contribute substantially to the rise of spatial audio as we know it today. In 1976 Dolby Laboratories took the idea of 4-2-4 channel matrixing and instead applied it to motion picture sound (Davis, 2003). Instead of following a symmetrical speaker arrangement, Dolby used three channels in front and a single surround channel (Torick, 1998). The investment in this technology, particularly from Dolby, would “prove to be the gateway through which consumer audio evolved from stereo to surround” (Davis, 2003). As storage capacity increased, it became feasible to use discrete rather than matrixed channels, and the number of output channels increased: the 1978 release of *Superman* marked the first use of a 5.1 channel soundtrack with a motion picture (Allen, 1991). Dolby continued to lead in the expansion of multichannel audio, providing surround encoding formats for both theatrical and home settings. More recently, even larger numbers of discrete channels have been proposed and implemented to varying degrees, including 7.1, 10.2, and 22.2 channels (Davis, 2003; Hamasaki, 2005). Mark Davis, an engineer

at Dolby, gives a thorough history of the development of spatial coding formats through the 20th century (2003).

Outside of the commercial realm (that is, mostly within academia) other more generalizable multichannel formats have been put forward that have found success within certain niches.

Chiefly among these are Ville Pulkki's Vector Base Amplitude Panning or VBAP (1997), and Distance-Based Amplitude Panning (DBAP), proposed independently by Lossius & Pascal Baltazar (2009) as well as Kostadinov & Reiss (Kostadinov, Reiss, & Mladenov, 2010). VBAP applies Blumlein's amplitude panning in full 3D, forming a vector basis composed of each loudspeaker in an enclosing array, allowing highly accurate spatial gestures. DBAP, though less rigorous, is more flexible and does not constrain the arrangement of speakers or listeners, making it popular for use in sound installations.

Wave Field Methods

While VBAP and DBAP can achieve convincing spatial effects under certain conditions, both of these methods fundamentally encode audio output in relation to a specific arrangement of loudspeakers, and are thus classified as *multichannel* methods. We will reserve the term *wave field* methods for those spatial formats, which seek to encode an entire sound field, independent of the arrangement of output transducers. This is done via Huygen's Principle, which states that each point on a progressing wavefront may instead be considered as a separate source, and its frequency-domain formulation, the Kirchoff-Helmholtz integral (Berkhout, de Vries, & Vogel, 1991). Wave front methods tend to be broadly separable into 1) Ambisonics, which is concerned with reproducing the incoming sound field around the listener, and 2) wave field synthesis,

which is concerned with reproducing the outgoing sound field emitted by one or more acoustic sources.

Ambisonics

Ambisonics came into being as another gem from the ruins of quadrophonic sound: in 1973 a mathematician and audio enthusiast named Michael Gerzon (1945-1996) put forward an encoding scheme that stood out from the crowd of competing matrixing schemes (Davis, 2003). Gerzon's scheme, later named Ambisonics, called for the use of spherical harmonic basis functions that could encode the portions of a sound field originating from many different directions around a listener's position (Gerzon, 1973). Though Gerzon worked out this system for an arbitrary number of bases (nth-order Ambisonics), the famous Sound Field Microphone Gerzon would later help develop was limited to first-order Ambisonics, including an omnidirectional signal and three orthogonal dipole terms (Gerzon, 1975). Lower orders of Ambisonics constitute a more severe truncation of the spherical harmonic decomposition and yield less spatial precision in reproduction. Since Gerzon's initial work, the increasing miniaturization of audio technology has allowed the development of 32-channel (Manola, Genovese, & Farina, 2012) and 64-channel (O'Donovan & Duraiswami, 2010) Ambisonic microphones, as well as software tools for spatialization using Higher-Order Ambisonics (Malham, 1999; Kronlachner, 2014). Since Ambisonic encoding is independent of any specific playback system, today's renewed interest in 3D sound has led to calls for using Ambisonics as a flexible production and distribution format for 3D audio content (Frank, Zotter, & Sontacchi, 2015).

Wave Field Synthesis

Wave field synthesis uses the same principle as Ambisonics to achieve the opposite aim: given an infinite amount of microphones around an acoustic source and an infinite number of loudspeakers in the same arrangement, each driven by the signal from its respective microphone, the wave fields in both cases should be identical. As mentioned earlier, Ader's Theatrophone was arguably a very simple implementation of this idea, and Fletcher and Stokowski's early three-channel orchestral transmissions were a better example of one-dimensional, highly truncated wave field synthesis. The basic theory behind wave field synthesis was outlined by William Snow (1955), and it received its modern formulation by Berkhout et al. (1993). In practice when a finite number of transducers is used, this leads to spatial aliasing above the spatial Nyquist frequency, which is around 1.7 kHz for practical transducer arrays (Berkhout et al., 1993).

Many methods have since been employed to reduce these shortcomings and allow practical implementations of wave field synthesis in two or three dimensions (Spors, Rabenstein, & Ahrens, 2008).

Back to Binaural

Binaural methods are the simplest 3D audio technology conceptually, yet they contain subtle difficulties and problems that still await a solution. It is because of this tension that binaural audio was the earliest 3D sound technique to be explored yet is only beginning to reach maturity. Though the late 20th century showed improvements in manikins and recording equipment (Paul, 2009), the dummy heads being used today are not categorically different from the binaural transmissions made by Fletcher in the 1930s. While they more accurately model a single averaged HRTF through the dummy's pinnae, this HRTF will inevitably deviate from that of the end listener, leading to degradations of front-back and up-down perception (Wenzel, Arruda,

Kistler, & Wightman, 1993) as well as sound source externalization over headphones (Hartmann & Wittenberg, 1996). Because measuring an individualized HRTF has traditionally been a time consuming and expensive process, much work has been done to quickly obtain individualized HRTFs for end-users of 3D audio simulations. Current approaches to this include wave equation simulations (Katz, 2001; Meshram, Mehra, & Manocha, 2014), database matching techniques (Andreopoulou, 2013), and reciprocal HRTF measurements, in which an emitter rather than a receiver is placed in the subject's ear canal (Zotkin, Duraiswami, Grassi, & Gumerov, 2006).

Transaural

Another route by which binaural content may be delivered bypasses headphones completely: the *transaural* technique uses crosstalk cancellation to deliver binaural content directly to a single listener over stereo loudspeakers. This requires using each channel of the stereo system to send delayed, phase-inverted versions of the sound coming from the other speaker, cancelling the acoustic 'crosstalk' that occurs when the left ear hears the signal from the right speaker and vice versa (Schroeder, Gottlob, & Siebrasse, 1974). The first crosstalk cancellation system was invented by Atal & Schroeder (1962) for the purposes of making instant A/B comparisons of different concert halls' acoustics to study listener preferences for music performance. The system was somewhat unstable as slight head movements caused it to induce severe spectral coloration upon the signals being reproduced. In 1985 Kendall and others used transaural properties to encode the earliest 3D sound broadcasts for CBS's *The Twilight Zone* (Gendel, 1985; Wolf, 1986; Kendall, 2015). Later improvements by Bauck & Cooper (1996) and Choueiri (2008) addressed these issues, and today it is possible to achieve spectrally uncolored transaural reproduction while maintaining sufficient interaural level difference for convincing 3D sound effects without the externalization problems that often accompany headphone listening. It must

be noticed that the various systems mentioned above are not mutually exclusive – a single 3D audio system might use Ambisonic content as the basis for binaural synthesis using a listener’s HRTF, but play back the content over a transaural system, creating a synergy out of the strengths of these different technologies to achieve the most convincing 3D sound possible.

Technology and Spatial Music

Art Music

The advent of electroacoustic technology – microphones, recordings, amplifiers, and loudspeakers – had a huge and far-reaching impact, not only on the frequency and temporal content of music composition, but also on the ways in which music could be spatialized. John Cage (1912-1992), another inheritor of Ives’s American Experimental tradition, quickly saw the potential of phonograph recordings and wireless radios, which he employed with spatial separation respectively in his *Imaginary Landscape Nos. 1 and 4* from 1939 to 1951, or the later installation “Writings through the Essay: On the Duty of Civil Disobedience” (1987). Both Cage and another aleatoric composer, Morton Feldman, would later make use of multiple tape recorders, routing each audio tape to a separate spatialized loudspeaker (Zvonar, 2006).

In contrast to Cage’s and Feldman’s use of space to embrace chaos, the European electroacoustic tradition began to see space as another parameter that could be serialized within highly deterministic musical structures. Pierre Schaeffer and Pierre Henry constructed a tetrahedral four-loudspeaker playback system for their *musique concrete*, which employed a potentiometer allowing them to route different channels of audio to specific speakers (Zvonar, 2006). Meanwhile, in Germany Karlheinz Stockhausen arguably began the modern electroacoustic tradition with his landmark *Gesang der Jünglinge*, which employed five loudspeakers around the

audience, and included spatial direction among the many parameters that Stockhausen rigidly serialized in the piece (Stone, 1963). Indeed, Stockhausen, in opposition to Brant's spatial philosophy, believed that distance effects should not be employed because they affected sound timbre. For this reason, Stockhausen believed that only sound direction was "the only spatial feature of sound worthy of compositional attention because it could be serialized" (Harley, 1997). Despite Stockhausen's large compositional influence, his ideas about space were not widely adopted, partially because the high tide of serialism began to ebb, and also because it was later understood that sound direction is also indelibly tied to sound timbre through the spectral filtering of the listener's HRTF.

After these initial explorations of multichannel spatial exploration, more ambitious playback environments were built. Probably the most famous example is the Phillips Pavilion at the 1958 World's Fair, designed by Iannis Xenakis (1922-2001), which hosted the tape piece *Poeme Electronique* by Edgard Varèse (1883-1965). Varèse recorded the piece on four separate tape recorders, which gradually desynchronized over time due to differences in playing speeds (Kendall, 2006). The final piece was presented over 425 loudspeakers in the finished pavilion, featuring nine different predetermined 'routes' for the sound to travel over (Zvonar, 2006). After this landmark installation, more ambitious and more flexible facilities were constructed, including those at IRCAM, the University of California at San Diego (UCSD), and Stanford University (Forsyth, 1985; Zvonar, 2006). The composer Roger Reynolds (1934-) explored spatial arrangements at UCSD, beginning with quadrophonic playback and pushing forward to 6- and 8-channel works (Zvonar, 2006). At Stanford, John Chowning created software to control the motion of a sound recording over a multichannel loudspeaker array, including a simulated Doppler shift using frequency modulation (Chowning, 1977). As personal computers and

multichannel sound cards proliferated, many of the different spatialization techniques described earlier were adopted by composers and sound installation artists. The exploration and understanding of spatial music is a continuing area of music research: composer and theorist Denis Smalley has put forth the idea that since acousmatic musical motion “implies a direction” (Smalley, 1986), that, “rather than being the final frontier of investigation, space should now move to centre stage to become the focal point of analysis” (Smalley, 2007).

Popular Music

It is well known that the Beatles were influenced by Stockhausen, as seen by his face’s inclusion on the cover of *Sergeant Pepper’s Lonely Hearts Club Band* (Richardson, 2015). Yet this did not extend to sound spatialization – in 1967 the Beatles were not even present for the stereo mixes of *Sergeant Pepper*, which were conducted far more quickly than the mono mixes. George Harrison reported that using two speakers seemed unnecessary and even made the music sound “naked” (Komara, 2010). Economic growth and the proliferation of stereo broadcasting and reproduction equipment would eventually make stereo playback the norm in pop music. Composer/producer Brian Eno sought to embrace and enhance the natural spatial characteristics of playback environments and landscapes with his ambient albums such as “Music for Airports” (1978) and “On Land” (1982), though these albums used only two-channel stereo for reproduction. Experimental/psychedelic band The Flaming Lips would go farther, performing early experiments in a parking lot full of their fans playing different tapes of prerecorded content through car stereos. This would later lead to their adventurous, though commercially unsuccessful, album *Zaireeka* (1997), which was released as four separate compact disks that had to be synchronously played on separate CD players. This distribution format made *Zaireeka* communal, almost like a concert recital, as at least four people had to be present to listen to the

album. Slight asynchronies in disk reading speeds, as with Varese's tape recorders used for *Poeme Electronique*, made every experience of *Zaireeka* different, giving the album a small but dedicated fan base.

The lack of a compact format for multichannel audio led to the dual releases in the late 1990s of the Super Audio CD format from Phillips/Sony, and the DVD Audio format from JVC, which featured 6- and 8-channel encoding respectively (Verbakel et al., 1998; Funasaka & Suzuki, 1997). Indeed, the Flaming Lips' next album, *Yoshimi Battles the Pink Robots* (2003), included a DVD version in Dolby Digital 5.1, taking advantage of the growing market in home theater systems to allow a single listener to hear immersive content from her living room (Rickert & Salvo, 2006). However, both formats failed to catch on outside the audiophile market, and by 2007 *The Guardian* called DVD Audio 'extinct' and Super Audio CDs 'dying' (Schofield, 2007). Today, larger multichannel formats for home and theatrical reproduction are multiplying, including object-based systems such as Dolby Atmos and DTS:X, as well as more traditional channel-based formats like Auro 3D (Dolby, 2014; Claypool, n.d.; Fonseca, 2015).

Conclusions and Thoughts for the Future

The experience and use of 3D sound in human culture has always been tied to the technological capabilities of the generation hearing it: for most of human history, this technology was limited to architectural spaces and music composition. Since the 19th century, more advanced technologies have allowed both the more accurate representation of real-world soundscapes as well as sonic spaces that have no correlate in physical reality. The former trend is now being

used in the rapid development of audio for virtual reality applications, which use head-mounted displays and headphones and are thus reliant on binaural reproduction methods (Begault, 2000; Xie, 2013). The latter trend of exploring new nonphysical spatial audio has been manifested in augmented reality systems, which use spatial auditory content to represent information beyond that encountered in the real world. Though these systems may use multichannel or wave field loudspeaker methods within certain controlled environments (Boren, Musick, Grossman, & Roginska, 2014), they also require binaural reproduction for deployment in day-to-day life, especially given the ubiquity of in-ear headphones or ‘earbuds’ in modern society (Sundareswaran et al., 2003). Multichannel methods persist for film audio and home theater systems, but in the future these markets may also face increased competition from binaural or transaural content as HRTF individualization methods improve.

What can the history of 3D sound tell us about the future of this field? Specific predictions are difficult because history shows us many contrasting trends that may lead to either rapid progress and standardization or else atomization and stagnation. Speculative investment without patient development may lead to another failure like quadrophonic sound, which lost the attention of the wider public and arguably delayed the implementation of more ambitious spatial audio formats. Perhaps some poor soul will outline the next century’s worth of progress but languish in obscurity, as did Blumlein. On the other hand, co-operation between scientists and content creators, as with the example of Fletcher and Stokowski, may lead to faster development than otherwise would have been expected. A hopeful sign is the recent collaboration between many different audio research institutions to create the Spatially Oriented Format for Acoustics (SOFA), a standardized file format that allows for the consolidation of many years’ and ears’ worth of HRTF research from around the world (Majdak et al., 2013). If this file format is

adopted by musicians, game designers, sound engineers, and other audio content creators, it might usher in a new renaissance in the use and experience of 3D sound. History contains many examples that elicit pessimistic predictions for the future. Thus it is best to continue to hope for the optimistic vision, yet those of us in the field, whether scientists or artists, must also continue working to make that vision a reality.

References

- Abel, J., Rick, J., Huang, P., Kolar, M., Smith, J., & J. Chowning. (2008). On the Acoustics of the Underground Galleries of Ancient Chavin de Huantar, Peru. In *Acoustics '08*. Paris.
- Allen, I. (1991). Matching the Sound to the Picture. In *Proceedings of the 9th Audio Engineering Society International Conference* (pp. 177–186). Detroit, Michigan.
- Amet, E. H. (1911). Method of and means for localizing sound reproduction. US Patent 1,124,580.
- Andreopoulou, A. (2013). *Head-Related Transfer Function Database Matching Based on Sparse Impulse Response Measurements*. Doctoral Dissertation, New York University.
- Arnold, D. (1959). The Significance of “Cori Spezzati.” *Music & Letters*, 40(1), 4–14.
- Atal, B. S., & Schroeder, M. R. (1962). Apparent sound source translator. US Patent 3,236,949.
- Bagenal, H. (1930). Bach’s Music and Church Acoustics. *Music & Letters*, 11(2), 146–155.
- Bagenal, H. (1951). Musical Taste and Concert Hall Design. *Proceedings of the Royal Musical Association*, 78(1), 11–29.
- Bauck, J., & Cooper, D. H. (1996). Generalized Transaural Stereo and Applications. *Journal of the Audio Engineering Society*, 44(9), 683–705.
- Begault, D. R. (2000). *3-D Sound for Virtual Reality and Multimedia*. Moffett Field, CA: National Aeronautics and Space Administration.
- Berkhout, A. J., de Vries, D., & Vogel, P. (1993). Acoustic control by wave field synthesis. *The Journal of the Acoustical Society of America*, 93(5), 2764–2778.
- Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization* (3rd ed.). Cambridge, MA: The MIT Press.
- Bloom, P. (1998). *The Life of Berlioz*. Cambridge, UK: Cambridge University Press.
- Blumlein, A. D. (1931). Improvements in and relating to Sound-transmission, Sound-recording, and Soundreproducing Systems. Great Britain Patent 394,325.
- Boren, B. B., & Longair, M. (2011). A Method for Acoustic Modeling of Past Soundscapes. In *Proceedings of the Acoustics of Ancient Theatres Conference*. Patras, Greece.
- Boren, B. B., Longair, M., & Orłowski, R. (2013). Acoustic Simulation of Renaissance Venetian Churches. *Acoustics in Practice*, 1(2), 17–28.
- Boren, B., Musick, M., Grossman, J., & Roginska, A. (2014). I HEAR NY4D : Hybrid Acoustic and Augmented Auditory Display for Urban Soundscapes. In *Proceedings of the 20th International Conference on Auditory Display (ICAD)*. New York, NY.
- Brant, H. (1978). Space as an Essential Aspect of Musical Composition. In E. Schwartz & B. Childs (Eds.), *Contemporary Composers on Contemporary Music* (pp. 223–242). New York: Da Capo Press.

- Broderick, A. E. (2012). *Grand Messe Des Morts: Hector Berlioz's Romantic Interpretation of the Roman Catholic Requiem Tradition*. Master's Thesis, Bowling Green State University.
- Burkholder, J. P., Grout, D. J., & Palisca, C. V. (2006). *A History of Western Music* (7th ed.). New York, NY: W. W. Norton and Company.
- Choueiri, E. (2008). Optimal crosstalk cancellation for binaural audio with two loudspeakers. Princeton University. <https://www.princeton.edu/3D3A/Publications/BACCHPaperV4d.pdf>, Accessed 6/23/2015, at 5:21 pm.
- Chourmouziadou, K., & Kang, J. (2008). Acoustic evolution of ancient Greek and Roman theatres. *Applied Acoustics*, 69(6), 514–529.
- Chowning, J. M. (1977). Simulation of Moving Sound Sources. *Computer Music Journal*, 1(3), 48–52.
- Claypool, B., Van Baelen, W., & Van Daele, B. (n.d.). *Auro 11.1 versus object-based sound in 3D*. Retrieved from <http://www.barco.com.cn/~media/Downloads/White papers/2012/WhitePaperAuro 111 versus objectbased sound in 3Dpdf.pdf>, Accessed 1/12/16, at 3:03 pm.
- Collins, P. (2008). Theatrophone: the 19th-century iPod. *New Scientist*, January 12, 44–45.
- Davis, M. F. (2003). History of Spatial Coding. *Journal of the Audio Engineering Society*, 51(6), 554–569.
- Declercq, N. F., & Dekeyser, C. S. A. (2007). Acoustic diffraction effects at the Hellenistic amphitheater of Epidaurus: seat rows responsible for the marvelous acoustics. *The Journal of the Acoustical Society of America*, 121(4), 2011–2022.
- Dixon, G. (1979). The Origins of the Roman “Colossal Baroque.” *Proceedings of the Royal Musical Association*, 106(1), 115–128.
- Dolby Laboratories (2014). *Authoring for Dolby Atmos Cinema Sound Manual*. San Francisco, CA. <http://www.dolby.com/us/en/technologies/dolby-atmos/authoring-for-dolby-atmos-cinema-sound-manual.pdf>, Accessed 1/12/16, at 2:38 pm.
- Eno, B. (1978). Liner Notes, “Ambient 1: Music for Airports.” <http://www.iub.edu/~audioweb/T369/enoambient.pdf>, Accessed 6/25/2015, at 3:25 pm.
- Eno, B. (1982). Liner Notes, “Ambient 4: On Land.” <http://www.iub.edu/~audioweb/T369/eno-ambient.pdf>, Accessed 6/25/2015, at 3:25 pm.
- Farnetani, A., Prodi, N., & Pompoli, R. (2008). On the acoustics of ancient Greek and Roman theaters. *The Journal of the Acoustical Society of America*, 124(3), 1557–1567.
- Fenlon, I. (2006). The Performance of cori spezzati in San Marco. In D. Howard & L. Moretti (Eds.), *Architettura e Musica Nella Venezia Del Rinascimento*. Bruno Mondadori.
- Fletcher, H. (1933). An Acoustic Illusion Telephonically Achieved. *Bell Laboratories Record*, 11(10), 286–289.
- Fletcher, H., & Sivian, L. J. (1927). Binaural telephone system. US Patent 1,624,486.
- Fonseca, N. (2015). Hybrid Channel-Object Approach for Cinema Post-Production using Particle Systems. In *Proceedings of the 139th Audio Engineering Convention*. New York, NY.

- Forsyth, M. (1985). *Buildings for Music: The Architect, the Musician, and the Listener from the Seventeenth Century to the Present Day*. Cambridge, MA: The MIT Press.
- Frank, M., Zotter, F., & Sontacchi, A. (2015). Producing 3D Audio in Ambisonics. In *Proceedings of the 57th AES International Conference*. Hollywood, CA.
- Fuller, S. (1981). Theoretical Foundations of Early Organum Theory. *Acta Musicologica*, 53(1), 52–84.
- Funasaka, E., & Suzuki, H. (1997). DVD-Audio Format. In *Proceedings of the 103rd Audio Engineering Society Convention*. New York, NY.
- Gendel, M. (1985, September 24). Hearing Is Believing On New “Twilight Zone.” *Los Angeles Times*.
http://articles.latimes.com/1985-09-24/entertainment/ca-18781_1_twilight-zone, Accessed 6/23/2015, at 5:40 pm.
- Gerzon, M. A. (1973). Periphony: With-Height Sound Reproduction. *Journal of the Audio Engineering Society*, 21(1), 2–10.
- Gerzon, M. A. (1975). The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound. In *Proceedings of the 50th Audio Engineering Society Convention*. London, UK.
- Hamasaki, K., Hiyama, K., & Okumura, R. (2005). The 22.2 Multichannel Sound System and Its Application. In *Proceedings of the 118th Audio Engineering Society Convention*. Barcelona, Spain.
- Harley, M. A. (1997). An American in Space: Henry Brant’s “Spatial Music.” *American Music*, 15(1), 70–92.
- Harmon, T. (1970). The Performance of Mozart’s Church Sonatas. *Music & Letters*, 51(1), 51–60.
- Hartmann, W. M., & Wittenberg, a. (1996). On the externalization of sound images. *The Journal of the Acoustical Society of America*, 99(6), 3678–3688.
- Hintermaier, E. (1975). The Missa Salisburgensis. *The Musical Times*, 116(1593), 965–966.
- Holman, P. (1994). Mystery Man. Peter Holman Celebrates the 350th Anniversary of the Birth of Heinrich Biber. *The Musical Times*, 135(1817), 437–441.
- Hospitalier, E. (1881). The Telephone at the Paris Opera. *Scientific American*, 45, 422–423.
<http://earlyradiohistory.us/1881opr.htm>, Accessed 6/17/2015, at 3:54 pm.
- Howard, D., & Moretti, L. (2010). *Sound and Space in Renaissance Venice*. Yale University Press.
- Huscher, P. (2006). Program Notes: Wolfgang Mozart, Notturmo in D, K.286, Chicago Symphony Orchestra.
https://cso.org/uploadedFiles/1_Tickets_and_Events/Program_Notes/ProgramNotes_Mozart_Notturmo.pdf,
 Accessed 6/15/2015, at 6:00 pm.
- Jahn, R. G. (1996). Acoustical resonances of assorted ancient structures. *The Journal of the Acoustical Society of America*, 99(2), 649–658.
- Katz, B. F. G. (2001). Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *The Journal of the Acoustical Society of America*, 110(5), 2440.

- Kendall, G. (1995a). A 3-D Sound Primer: Directional Hearing and Stereo Reproduction. *Computer Music Journal*, 19(4), 23–46.
- Kendall, G. (1995b). The Decorrelation of Audio Signals and Its Impact on Spatial Imagery. *Computer Music Journal*, 19(4), pp. 71–87.
- Kendall, G. S. (2006). Juxtaposition and Non-motion: Varèse bridges early modernism to electroacoustic music. *Organised Sound*, 11(2), 159–171.
- Kendall, G. (2015). Personal communication.
- Komara, E. (2010). The Beatles in Mono: The Complete Mono Recordings. *ASRC Journal*, 41(2), 318–323.
- Kostadinov, D., Reiss, J. D., & Mladenov, V. (2010). Evaluation of Distance Based Amplitude Panning for Spatial Audio. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal (ICASSP)*. Dallas, TX.
- Kronlachner, M. (2014). *Spatial Transformations for the Alteration of Ambisonic Recordings*. Master's Thesis: University of Music and Performing Arts, Graz, Austria.
- T, H. W. (1930). The Use of Sound in Navigation. *The Journal of the Acoustical Society of America*, 1(3), 403–409.
- Lindsay, B. (1966). The Story of Acoustics. *Journal of the Acoustical Society of America*, 39(4), 629–644.
- Lossius, T., & Pascal Baltazar, T. (2009). DBAP - Distance-Based Amplitude Panning. In *Proceedings of International Computer Music Conference (ICMC)*. Montreal, Quebec.
- Lubman, D., & Kiser, B. H. (2001). The History of Western Civilization Told Through the Acoustics of its Worship Spaces. In *Proceedings of the 17th International Congress on Acoustics*. Rome, Italy.
- MacDonald, C. (2009). Scoring the Work: Documenting Practice and Performance in Variable Media Art. *Leonardo*, 42(1), 59–63.
- MacGowan, K. (1957). Screen Wonders of the Past: And to Come? *The Quarterly of Film Radio and Television*, 11(4), 381–393.
- Majdak, P., Iwaya, Y., Carpentier, T., Nicol, R., Parmentier, M., Roginska, A., ... Noisternig, M. (2013). Spatially Oriented Format for Acoustics. In *Proceedings of the 134th Audio Engineering Society Convention*. Rome, Italy.
- Malham, D. G., & Anthony, M. (1995). 3-D Sound Spatialization using Ambisonic Techniques. *Computer Music Journal*, 19(4), 58–70.
- Malham, D. G. (1999). Higher order Ambisonic systems for the spatialisation of sound. In *Proceedings of the International Computer Music Conference* (pp. 484–487). Beijing, China.
- Manola, F., Genovese, A., & Farina, A. (2012). A comparison of different surround sound recording and reproduction techniques based on the use of a 32 capsules microphone array, including the influence of panoramic video. In *Audio Engineering Society 25th UK Conference: Spatial Audio in Today's 3D World*. York, UK.

- Meshram, A., Mehra, R., & Manocha, D. (2014). Efficient HRTF Computation using Adaptive Rectangular Decomposition. In *AES 55th International Conference*. Helsinki, Finland.
- McDonald, M. (2004). Silent Narration? Elements of Narrative in Ives's The Unanswered Question. *19th-Century Music*, 27(3), 263–286.
- Moretti, L. (2004). Architectural Spaces for Music: Jacopo Sansovino and Adrian Willaert at St Mark's. *Early Music History*, 23(2004), 153–184.
- Navarro, J., Sendra, J. J., & Muñoz, S. (2009). The Western Latin church as a place for music and preaching: An acoustic assessment. *Applied Acoustics*, 70(6), 781–789.
- O'Donovan, A., & Duraiswami, R. (2010). Audio-Visual Panoramas and Spherical Audio Analysis using the Audio Camera. In *Proceedings of the 16th International Conference on Auditory Display (ICAD2010)* (pp. 167–168). Washington, DC.
- Paul, S. (2009). Binaural Recording Technology: A Historical Review and Possible Future Developments. *Acta Acustica United With Acustica*, 95, 767–788.
- Prasad, M. G., & Rajavel, B. (2013). Acoustics of Chants, Conch-Shells, Bells and Gongs in Hindu Worship Spaces. In *Acoustics 2013* (pp. 137–152). New Delhi, India.
- Pulkki, V. (1997). Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *Journal of the Audio Engineering Society*, 45(6), 456–466.
- Reznikoff, I. (2008). Sound resonance in prehistoric times: A study of Paleolithic painted caves and rocks. In *Acoustics '08* (pp. 4137–4141). Paris.
- Richardson, C. E. (2015). *Stockhausen's influence on popular music: an overview and a case study on Björk's Medúlla*. Master's Thesis, Texas State University.
- Rickert, T., & Salvo, M. (2006). The distributed Gesamtkunstwerk: Sound, worlding, and new media culture. *Computers and Composition*, 23(3), 296–316.
- Rindel, J. H. (2011a). The ERATO Project and Its Contribution to Our Understanding of the Acoustics of Ancient Theatres. In *The Acoustics of Ancient Theatres Conference*. Patras, Greece.
- Rindel, J. H. (2011b). Echo Problems in Ancient Theatres and a Comment to the “Sounding Vessels” Described by Vitruvius. In *The Acoustics of Ancient Theatres Conference*. Patras, Greece.
- Ritchey, M. (2010). Echoes of the Guillotine: Berlioz and the French Fantastic. *19th-Century Music*, 34(2), 168–185.
- Rosenthal, K. A., & Mendel, A. (1941). Mozart's Sacramental Litanies and Their Forerunners. *The Musical Quarterly*, 27(4), 433–455.
- Schofield, J. (2007). No taste for high-quality audio. *The Guardian*.
<http://www.theguardian.com/technology/2007/aug/02/guardianweeklytechnologysection.digitalmusic>,
 Accessed 1/12/16, 2:02 pm.

- Schroeder, M. R., Gottlob, D., & Siebrasse, K. F. (1974). Comparative study of European concert halls: correlation of subjective preference with geometric and acoustic parameters. *Journal of the Acoustical Society of America*, 56(4), 1195–1201.
- Slotki, I. W. (1936). Antiphony in Ancient Hebrew Poetry. *The Jewish Quarterly Review*, 26(3), 199–219.
- Smalley, D. 1986. Spectro-morphology and structuring processes. In S. Emmerson (ed.) *The Language of Electroacoustic Music*, pp. 61–93. New York: Harwood Academic.
- Smalley, D. (2007). Space-form and the acousmatic image. *Organised Sound*, 12(1), 35–58.
- Snow, W. (1955). Basic principles of stereophonic sound. *IRE Transactions on Audio*, 3(2), 42–53.
- Soeta, Y., Shimokura, R., Kim, Y. H., Ohsawa, T., & Ito, K. (2013). Measurement of acoustic characteristics of Japanese Buddhist temples in relation to sound source location and direction. *The Journal of the Acoustical Society of America*, 133(5), 2699–2710.
- Spors, S., Rabenstein, R., & Ahrens, J. (2008). The Theory of Wave Field Synthesis Revisited. In *Proceedings of the 124th Audio Engineering Society Convention*. Amsterdam, The Netherlands.
- Stevens, D. (1982). A song of fortie parts, made by MR. Tallys. *Early Music*, 10(2), 171–182.
- Stone, K. (1963). Karlheinz Stockhausen: Gesang der Junglinge (1955/56). *The Musical Quarterly*, 49(4), 551–554.
- Strutt, J. W. (1875). On Our Perception of the Direction of a Source of Sound. *Proceedings of the Musical Association*.
- Sunier, J. (1986). A History of Binaural Sound. *Audio*, 36–44.
- Sundareswaran, V., Wang, K., Chen, S., Behringer, R., McGee, J., Tam, C., & Zahorik, P. (2003, October). 3D audio augmented reality: implementation and experiments. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality* (p. 296). IEEE Computer Society.
- Torick, E. (1998). Highlights in the history of multichannel sound. *Journal of the Audio Engineering Society*, 372, 368–372.
- Verbakel, J., van de Kerkhof, L., Maeda, M., & Inazawa, Y. (1998). Super Audio CD Format. In *Proceedings of the 104th Audio Engineering Society Convention*. Amsterdam, The Netherlands.
- Vitruvius, M. (1914). *Vitruvius: The Ten Books on Architecture*. (M. Morgan, Ed.). Cambridge, MA: Harvard University Press.
- Wenzel, E., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America*, 94(1), 111–123.
- Wolf, R. (1986, April 18). At Northwestern, They're Reshaping World Of Sound. *Chicago Tribune*.
http://articles.chicagotribune.com/1986-04-18/entertainment/8601280476_1_spatial-sound-outer-ear,
 Accessed 6/26/2015, at 4:28 pm.
- Xie, B. (2013). *Head-related transfer function and virtual auditory display* (2nd ed.). J Ross.

Zotkin, D. N., Duraiswami, R., Grassi, E., & Gumerov, N. a. (2006). Fast head-related transfer function measurement via reciprocity. *The Journal of the Acoustical Society of America*, 120(4), 2202–2215.

Zvonar, R. (2006). A History of Spatial Music. *eContact*, 7(4).
http://cec.sonus.ca/econtact/7_4/zvonar_spatialmusic.html, Accessed 6/20/2015, at 1:28 pm.

Acknowledgments

Many thanks to Durand Begault, Gary Kendall, and Agnieszka Roginska, who provided many starting points for the exploration of this large subject. Thanks also to John Krane for introducing me to spatial audio many years ago through a *Zaireeka* listening party with four exceptionally synchronized play-button-pushers.