

## chapter four | sound

'Ladies and Gentlemen – isn't this a marvellous invention?'<sup>1</sup>

'Unlike a portrait, the reproduction of a dead voice gives one an uneasy feeling.'<sup>2</sup>

'A cinema patron and his wife have been driven from one of our few remaining places of public worship by an overdose of decibels ... It ranged from rock-pop to the factory floor, the wilful infliction of punk rubbish on our eardrums to the enforced hammerings and screechings of workshops.'<sup>3</sup>

Despite this customer's somewhat less than enthusiastic reception of the newly released Dolby stereo process, audio recording has been always been associated with moving image technology in some shape or form throughout the period of the latter's existence. Recordings have been produced for playing in synchronisation with the picture for all but the first three decades of commercial film (and even then some activity in this area took place), and for almost all publicly broadcast television. Synchronised sound is also an integral component of all video recording technology, of digital offline and of Internet-based moving image content. The evolution of these audio technologies will be considered alongside the video technologies with which they are associated in subsequent chapters. This discussion, therefore, is concerned primarily with the evolution of audio recording, manipulation and reproduction technologies specifically associated with film.

The overwhelming majority of historical research related to film sound technology tends to focus on the 'key moment' of the conversion process, which in North America and Europe took place roughly between 1926 and 1932. It is not difficult to understand why. At the outset of this process, hardly any feature-length films were produced and shown commercially with a synchronised audio recording. By the close of it, 'silent' films, as Scott Eyman elegantly puts it, 'belonged to the permanent, irremediable past'.<sup>4</sup> Three generations of writers and historians, therefore, have (correctly) identified this period as being one of great industrial and cultural change. So has the Hollywood film industry itself, mythologising the conver-

sion through productions such as *Hollywood Cavalcade* (1939, dir. Irving Cummings) and *Singin' in the Rain* (1952, dir. Stanley Donen & Gene Kelly), which celebrate the arrival of sound, and *Sunset Boulevard* (1950, dir. Billy Wilder), which condemns it. But many writers have fallen into the trap of reading huge, agenda-setting *ideological* developments into this process as well, when the real significance of the conversion as a 'key moment' lies in the convergence of a complex framework of technological, cultural and economic factors which enabled the mass rollout of synchronised sound to take place when it did. For example, Laura Mulvey suggests that Warner Bros.' collaboration with AT&T during the mid-1920s, which resulted in the Vitaphone process, enabled the 'negotiation of the technology into a showbusiness reality'.<sup>5</sup> But however valid her arguments are as to Warner Bros. having made the business case for sound *per se*, her essay completely ignores the fundamental technological flaws in Vitaphone which made it essentially incompatible with the film industry's industrial practices (and which were well known a decade earlier), the fact that it only remained a 'showbusiness reality' for about four years and that, in all likelihood, silent cinema would have returned if other, more suitable methods of sound recording and reproduction had not been introduced as the result of Warner Bros. making that business case.

The 1927–30 'conversion', therefore, gave rise to what is probably the most extreme example of the 'key moment' model of technological historiography, one which in this chapter I shall try to set in the context of developments before and since. In doing so, I will argue that the long-term impact of audio technology on film production and consumption is better understood through a more balanced approach to its ongoing development and implementation (or, in some cases, lack of implementation) throughout the late nineteenth and twentieth centuries. From this standpoint it is possible to extrapolate a clearer picture of the ways in which audio technology had affected the production and exhibition of films.

### The origins of audio recording and their links to moving image technology

Given that the following four decades would be spent trying to add synchronised, recorded sound to moving pictures, it is more than a little ironic that in all likelihood, the earliest successful moving image technology was invented for the sole purpose of adding illustration to sound recordings! One of Thomas Edison's patent applications relating to the Kinetograph stated that 'I am experimenting upon an instrument which does for the eye what the Phonograph does for the ear',<sup>6</sup> and his assistant W. K. L. Dickson claims to have successfully demonstrated a Phonograph mechanically synchronised to a projected film on 6 October 1889.<sup>7</sup>

Audio recording came first by a decade or so. On 7 December 1877 Edison demonstrated his 'Phonograph' to the public at the offices of the *Scientific American* magazine. The device, though crude, demonstrated the basic technical principles through which sound would be recorded until the mid-1920s, and would be reproduced domestically well into the 1950s. It consisted of a steel cylinder coated with tin foil, onto which a recording was made acoustically. As the cylinder was rotated

at a (more or less) constant speed by hand-cranking the shaft, the operator spoke into a diaphragm. This was mechanically linked to a stylus which inscribed a linear indentation (groove) into the constantly rotating and laterally moving (thus preventing the groove from being overwritten in the cylinder's following rotation) foil surface. As the noise level increased, the greater air pressure forced the needle deeper into the foil, while at lower volumes the indentation was lighter. Edison discovered that when, after recording, a similar needle was passed over the groove at the same constant speed but at much lower pressure, its vibrations could be reproduced as changes in air pressure created through a similar diaphragm, thereby playing back the recorded sound.

With the apocryphal words 'Mary had a little lamb',<sup>8</sup> Edison had successfully demonstrated the basic technical principles of analogue audio recording, and shown that they could be made to work in practice. As with film, the issue was then to identify how this technology could be commercially exploited. A decade later, his (or, more accurately, Dickson's) Kinetograph/Kinetoscope system for recording and reproducing film-based moving images would ultimately be eclipsed by minor variations to the technology introduced by Armat and Latham in the US and Lumière, Paul and Prestwich (among others) in Europe, in that projection before a large audience was what made the sums add up. The Kinetoscope's fatal flaw, in other words, was that it was an expensive piece of capital equipment which could only show films to one person at a time.

As with Edison's movie equipment, the Phonograph represented an impressive scientific feat, but one which, in the first instance, did not seem to have any obvious commercial application. Edison's original intention was to market the Phonograph as a device for consumers to create and replay their own recordings, similar to George Eastman's simple and mass-produced stills cameras. As David Morton explains:

Edison's original phonograph was merely a clever parrot, or better yet, an aural mirror. The phonograph's lacklustre sales soon made it clear that few Americans would be satisfied with simply recording themselves. Instead, buyers rewarded those who used the phonograph to create a system for mass-produced entertainment, purchasing millions of records (or later tapes or compact discs) for their personal enjoyment.<sup>9</sup>

The promoters of sound recording technology, it seems, took two to three decades to work out what the nascent movie industry (represented by the Lumière brothers) discovered in one fell swoop on the evening of 28 December 1895: that selling a single recording to multiple customers at the same time was the route to commercial success. Not that it is too difficult to find a justification for Edison's line of thought. The previous chapter discussed a number of colour film processes which produced impressive results on the screen, but which ultimately fell by the wayside because it was too difficult and/or expensive to produce multiple copies from a single original. Edison's Phonograph suffered from the same flaw. As a rapidly emerging market for the hardware and software needed to reproduce pre-recorded sound in the home emerged (which itself had a number of Victorian precedents, such as pianola rolls),

Edison's cylinder technology found itself unable to economically supply it because, as with reversal colour film processes, that was not what it was either intended or designed for.

Initially, every cylinder had to be individually recorded. It was not until 1912 – long after the film industry had perfected the means of producing release prints in the quantities needed to satisfy the exhibition market as it then existed – that Edison had developed a mass-duplication system for cylinders (the 'Blue Amberol' process), based on electroplating, which offered an acceptable quality of reproduction compared to the original. By this stage, Edison had pretty much been eclipsed in the audio market by a variant of the technology more suited to customer demands, just as he had been by the systematic emergence of film projection in Europe.

Acoustic recording using discs, as distinct from cylinders, as the carrier, emerged not long after Edison's initial experiments. In New York the German émigré Emile Berliner launched the first commercially marketed form of this technology, which established a rapidly growing market share during the 1890s. Although the basic technical principle of mechanical, acoustic recording used in disc technology was identical to that of Edison cylinders, Berliner managed to adapt it to facilitate the mass-production of release copies from a single original recording far more efficiently than was possible with Edison cylinders. The master recordings consisted of a rigid disc coated with wax, into which a stylus inscribed the groove in a spiral formation, starting at the outer edge and moving inwards. Variations in input volume registered as lateral movements in the wax rather than the depth of the indentation. Both of these changes meant that the wax original could be electroplated, a metal 'stamper' produced and release copies pressed on a highly durable medium far more cheaply and efficiently than with cylinders; and the records sold in shops offered a much higher sound quality than their cylinder equivalents.<sup>10</sup> By the mid-1900s, sales of pre-recorded music on discs had overtaken those of cylinders, although production of the latter struggled on into the early 1920s.

I have cited Edison as a major case study around the origins of industrial practice in audio recording because, by the mid-1900s, his experiences offer a clear demonstration of how the film and audio industries were heading in the direction of two separate economic models. Both were, by this stage, reliant on the means of producing multiple copies from a single original. But films were increasingly being consumed by mass audiences in a formal, auditorium setting; whereas most copies of audio recordings were sold outright to private individuals for playback in a domestic setting (the impact of the latter on film industry economics will be discussed in relation to home video in chapter six). Those economic models were greatly influenced by the strengths and weaknesses of the technologies themselves: characteristics which, with both moving images and audio, Edison's designs found themselves hopelessly fighting against. Despite a last-ditch rearguard action in the form of the MPPC 'adventure' of 1908–17 (see chapter two), which attempted to package Edison's essentially obsolete moving image technologies with the commercially successful Eastman film stock, he ended up paying the price of the pioneers on both fronts. By the mid-1920s the Edison Company had

disappeared, virtually without trace, from both the film and the recorded sound industries.

The reconvergence of those two economic models into a single mass-medium consisting of edited films shown in synchronisation with a recorded soundtrack took place gradually over the following two decades. It was driven largely by developments in audio technology which made it increasingly compatible with the existing infrastructure of film production, duplication/distribution and exhibition. The 1926–32 ‘key moment’ certainly marked a period of intense activity in the rollout of synchronised sound. But to overemphasise its overall importance would be to risk falling into the trap of assuming that synchronised film and sound did not exist on a commercial basis before then, and/or that research and development in the area suddenly stopped once the last cinema had been wired up. The cultural, commercial and technological developments in film sound which happened during the intervening period (very roughly, 1905–26), place the traditionally understood conversion stage into a somewhat more organic and progressive context. This stage can usefully be considered in two separate categories: live and recorded film sound.

### **Film sound before the conversion: live performance**

It has become a recurrent cliché among cinema historians that ‘there was no such thing as a silent film’: rather, there was a period during which the dominant cultural and economic mode of practice dictated that the film and the sound which went with it were produced and supplied separately. Although some early film exhibition, notably of newsreels and other non-fiction subjects, did take place in total silence, this would eventually become the exception which proved a soon to be established rule.<sup>11</sup> The ‘sound of silents’, therefore, was generally arranged by the exhibitor, and tended to consist of one or a combination of four forms of live performance: a spoken lecture or commentary; a live musical performance, ranging from a single instrumentalist to a live ensemble or orchestra; sound effects corresponding to, and generated in approximate synchronisation with, the action in the film; and, more rarely (in the European and North American film industries, at least), a theatrical performance by actors which took place near the screen and in view of the audience.

It is likely that lectures were the dominant mode of performance during the first decade of industrial-scale film production and exhibition. The reason for this is twofold. Firstly, this was already a well-established practice in ‘magic lantern’ slide exhibition, where sets of slides and printed lecture texts were supplied by the producing company; the latter were usually read verbatim. Secondly (and as we shall see in the following chapter), cinema exhibition was initially an itinerant business. Fairgrounds, music halls and other ‘multi-purpose’ venues were the usual locations, and therefore the equipment and infrastructure associated with film exhibition had to be easily portable. This militated against the use of acoustically recorded sound, live orchestras or performances which required the use of sets or lighting.

This period pre-dated the production/distribution/exhibition model which characterises today’s global film industry and which started to become established with

the emergence of Hollywood towards the end of World War One. Copies of films were sold outright by producers to exhibitors, and one of the key ways in which they advertised their products was through the use of extensive printed catalogues. These often contained elaborate descriptive prose relating to each title offered for sale, which in turn would form the basis for 'ad-libbed' lectures and commentaries performed by the exhibitor. As Richard Crangle points out, these were not usually as crucial to understanding the action in a film as lantern lectures were to the narratives in slide shows, and were therefore not usually scripted and performed verbatim.<sup>12</sup> But as the desire to convey more complex narratives through film (both fictional and non-fictional) developed during the 1900s, lecturers did sometimes find themselves in the position of having to deliver information to audiences which the embryonic grammar of film editing was then unable to. For example, the British crime drama *A Daring Daylight Burglary* (1903, dir. Frank Mottershaw), is described by one writer as 'remarkable for its precocious editing experiments':<sup>13</sup> so remarkable, in fact, that a crucial plot point (that one policeman has telephoned another, several miles away, to arrange for a criminal's apprehension upon his arrival), fails to be conveyed by the film itself at all. The audience was reliant on a commentator to provide this information (who in turn got it from the catalogue description), without which the action on the screen makes little sense.

Lectures and commentaries became less common and had effectively disappeared by the outbreak of World War One. Purpose-built cinemas became the norm, in which an orchestra, organ, sound effects equipment and eventually audio playback equipment could be permanently installed. The main precedent for film commentaries, the magic lantern, died out rapidly during the early 1910s as film replaced it as the dominant visual entertainment medium. And the films themselves changed. Editing became more sophisticated. Filmmakers learnt to convey specific information or meaning through the juxtaposition of different sorts of shot, and audiences learnt how to 'read' it. The intertitle (a short section of prose or dialogue, rarely longer than twenty words, shown momentarily against an opaque background in between shots) gradually replaced the commentator in situations such as the example given above. The length of films increased substantially, and the early 1910s saw the emergence of the 'feature' film, usually lasting significantly over an hour, as the main component in a cinema presentation. Sound, therefore, no longer had to fulfil a direct narrative function in the way that it did during the lecture and commentary period. The result was that a transition took place, at the end of which the live musical performance had emerged as the primary form of film sound in cinema exhibition.

As with the provision of lecture notes, the nature and extent to which the form and content of musical performances was determined at the performance stage, relative to the evolution of practices within individual cinemas, varied considerably from film to film and from exhibitor to exhibitor. By the 1920s the major Hollywood and European studios often provided full-scale orchestral scores, complete with detailed instructions as to the projection speed needed in various different sequences of the film (as has been noted in chapter two, camera and projection speeds were not standardised until the conversion to sound) in order to synchronise them to a per-

formance at a given tempo.<sup>14</sup> Performance in the cinema ranged from large orchestras in prestigious, city centre locations to an individual pianist or organist in suburban or rural second-run venues. Depending on the musical ability and preferences of individual performers, the accompaniments were frequently improvised. In addition to musicians, a variety of machines were marketed which produced specific sound effects, either mechanically or pneumatically. The 'Allfex', sold in Britain from 1910, claimed to produce over fifty effects, including a steam engine, rain, hail, smashing china, a machine gun and a fire alarm.<sup>15</sup> However, their reception by audiences was problematic, with many complaining of inappropriate use or over-use of these effects.<sup>16</sup> As with first generation synchronised sound (see below), it seems that their use had all but died out by the end of the 1910s.

More elaborate modes of live performance designed to accompany a film projection were much rarer in the West. Where they did occur they were usually the work of experimental or avant-garde filmmakers, and did not represent the mainstream form of the medium. For example, *Entr'acte* (1924, dir. René Clair) was made to be shown during the performance of a ballet by the avant-garde composer Erik Satie. In other film cultures they were far more widely used, however, most notably in India and Japan, where cinema initially served to augment other, established forms of narrative culture. For example, while the lecture/commentary mode was essentially a primitive device in Western cinema – a throwback from an earlier form of entertainment, reluctantly used only until film could stand on its own two feet – in Japan, it was a deeply-rooted form of cultural expression to which film was effectively regarded as an added extra, as the director Akira Kurosawa recalls:

The narrators not only recounted the plot of the films, they enhanced the emotional content by performing the voices and sound effects, and providing evocative descriptions of events and images on the screen – much like the narrators of the *bunraku* puppet theatre. The most popular narrators were stars in their own right, solely responsible for the patronage of a particular theatre.<sup>17</sup>

It is perhaps because North American and European filmmakers and technologists had a cultural view of cinema as a unique, self-contained medium which should not need to interact with other, established forms of entertainment in the way that it did in many Asian societies, that the push to develop audio recording and reproduction into a technology which was specifically suited for use in synchronisation with moving images originated primarily in the USA, and was based to a certain extent on European expertise. In other words, an underlying assumption existed to the effect that if the picture could be canned, so could the sound.

### Film sound before the conversion: recorded sound

In the section on Edison above, we have seen how the nascent audio industry discovered by a process of trial and error that the most economically successful way of

exploiting its technology was in selling mass-duplicated copies of music recordings for reproduction in the home. Equipment which enabled consumers to create their own recordings, or which facilitated their use in a business context, found virtually no market.<sup>18</sup> The film industry, too, evolved an economic model which was predicated on exploiting multiple copies of a single production, though a combination of technical (for example, the fire risk from nitrate film, the cost of equipment and the expertise needed to operate it) and cultural (initial film exhibition practices emerged from other forms of theatrical performance) factors dictated that consumption took place in a communal, rather than a domestic setting. Beginning at around the turn of the century and continuing well into World War One, there were a number of sustained attempts to combine the two technologies. While they enjoyed limited but notable popular success, they are of interest mainly because they illustrated clearly what had to be done in order to truly converge the technologies. These embryonic attempts to link moving images and audio recording, therefore, set the cultural agenda for a process of research and development which culminated in the 1920s, which made its first significant impact with the 'key moment' conversion process and which has essentially continued to the present day.

Edison's failure to establish a market for synchronised sound films in the 1890s did not mean that they disappeared from circulation for the following three decades. On the contrary, they found a niche market which exploited their capabilities and minimised the impact of their drawbacks, and which continued until World War One. The systems developed all used acoustic recording and amplification, mostly with the Berliner-type discs which had largely eclipsed Edison cylinders in the domestic audio market. The overwhelming majority of 'sound films' made were of musical or theatrical performances, ranging from popular songs to opera. In some ways they could be compared to modern music videos, and in the decade-and-a-half before feature films became the norm, their length, style and format was probably not radically different from those of other films which were produced and exhibited at the time.

In France, then the world's most economically advanced film industry, the company founded by Léon Gaumont began research in 1896, which culminated in the launch of its cylinder-based 'Phono-Cinéma-Théâtre' system at the Paris Exposition in 1900. This synchronised a musical accompaniment to footage of a number of well-known theatrical performers. It did not catch on with audiences, and the venture lost 150,000 Francs during the two months of the exhibition.<sup>19</sup> Four years later, in Germany, the movie pioneer Oskar Messter launched his disc-based 'Biophon' in 1903, and is said to have produced over 500 three-minute musical shorts which were shown successfully in Berlin during the following decade.<sup>20</sup> Britain was represented principally by Cecil Hepworth, whose 'Vivaphone' system of 1911 introduced a novelty: not only were musical numbers filmed and recorded, but he also produced footage of campaign speeches made by Lord Birkenhead and Andrew Bonar-Law, which must be among the earliest examples of political propaganda in synchronised film.<sup>21</sup> The following year, Gaumont was back; this time with the 'Chronophone' device, which mechanically interlocked a phonograph disc player and a film projector

to ensure synchronisation. Like its other European predecessors Chronophone was exported to the US on a limited scale, where sound films were also regularly shown during this period, either for novelty value or as a component of variety hall or theatrical performances.

All of these first generation sound systems shared five key characteristics which defined the extent of their use in conjunction with film. Firstly, acoustic recordings (both cylinders and discs) were what would now be termed a 'write once' medium: after recording, the signal characteristics could not be altered in any way. Secondly, no editing of the recording's content was possible, analogous to the way that films could be cut and spliced. Thirdly, they were time-limited to (approximately) two minutes in the case of cylinders, and three for discs. Fourthly, there was the issue of synchronisation in playback. As all these systems relied on two separate carriers for the picture and sound, a means was needed of regulating their playback speed. They ranged from sophisticated to effectively non-existent. Chronophone, for example, used a crude form of mechanical interlocking which enabled the projectionist to adjust the relative speeds of the projector and turntable. Vivaphone only enabled the projector speed to be adjusted in order to match that of the record, but did provide a visual indicator in the projection box which enabled these adjustments to be made with some degree of accuracy.<sup>22</sup> At the other end of the scale were systems which, to all intents and purposes, ran the records 'wild', with consequent catastrophic results for the accuracy of synchronisation.

And finally, the acoustic nature of the medium imposed severe limitations. Recording had to take place close to a large horn – so close, in fact, that it would be visible in the frame if any attempt were made to actually record the picture and sound simultaneously. Therefore, almost all of these 'synchronised' films were produced by a performer lip-synching in front of a camera while a copy of the record (which had been recorded earlier) was played. In the case of the American 'Cameraphone' and the British Waltham 'Singing Pictures' series, these were not even specially produced for the film, but rather were off-the-shelf commercial recordings that were already on sale to consumers.<sup>23</sup> The other restriction imposed by acoustic technology was that of amplification. As has been mentioned above, playback of acoustic recordings was accomplished by passing a needle through the groove, which vibrated in response to the indentations made by the recording needle. The changes in air pressure generated by these vibrations were amplified by resonance within a metal horn, thereby replaying the recording. However, the extent of this amplification was inherently limited by the width of the needle, which in turn was limited by the width of the groove (or, in an Edison cylinder, its depth). Acoustic reproduction offered no means of increasing the playback volume without practical limit, something which was essential if recorded sound in large auditoria was to become a regular fixture. A number of experiments were attempted in order to overcome the problem, most notably that of pneumatic amplification, in which the vibrations of the playback needle operated a valve which released large quantities of compressed air under high pressure into the playback horn, in proportion to its vibration in the groove. While they succeeded in increasing the capacity of acoustic

amplification, this was, according to many contemporary accounts, at the expense of sound quality (a subtle indication of this can be found in the acoustic method being described as the 'siren type' of amplifier by one technical historian).<sup>24</sup> The Chronophone playback system was succinctly – if a little unkindly – described by Eyman as 'jerry built'.<sup>25</sup>

In a last-gasp attempt to launch acoustic film sound as a viable commercial proposition, Edison briefly re-emerged on the scene in 1913 with the cylinder-based 'Kinetophone' system, which attempted – unsuccessfully – to address some of these issues. The mechanical synchronisation was more sophisticated than with any of its predecessors, the maximum running time was increased to approximately six minutes and the pneumatic amplification tweaked to improve the playback quality.<sup>26</sup> But after a six-month run at a New York theatre, acoustic recording – and Edison – disappeared from the film industry; this time, for good.

### Classical Hollywood and electric recording

In order to explain why the wholesale adoption of synchronised sound happened when it did, it is necessary to identify and analyse two factors: the technologies which were appropriated for the purpose, and the cultural and economic factors which precipitated their use in synchronisation with moving images. As has been mentioned above and in previous chapters, the (approximate) period of 1910 to 1917 witnessed a wide-ranging process of change in the North American and Western European film industries. Itinerant exhibition and temporary venues gave way to purpose-built cinema theatres; the practice of producers selling copies of films directly to exhibitors was replaced by the rental-based distribution system, very similar to the one in place today; the beginnings of the studio system were established; and Donald J. Bell's 'fondest hope' of technological standardisation started to take shape. The form and content of the films which were produced and shown changed, too. The industry's first decade-and-a-half was characterised by short films, usually based around real events, special effects, or simple and embryonic fictional narratives. By the early 1910s, the films were getting longer and the stories they told were becoming more complex. A self-contained 'language' of cinema began to emerge, based on a combination of acting, directing, visual and editing devices. This has been termed the 'continuity system', or 'classical Hollywood', and was developed largely in America during this period.<sup>27</sup> It was a language that proved straightforward for filmmakers to implement and easy for audiences to understand: as one cinemagoer noted in 1917, 'ninety-nine picture fans in every hundred can instantly tell whether the continuity in a picture is good or bad'.<sup>28</sup> With the benefit of hindsight, it is not at all difficult to understand why acoustic recording declined and disappeared during exactly the same period as the continuity system took hold: the technical limitations of the former were simply incompatible with the requirements of the latter.

In an essay on sound reproduction in cinemas, Rick Altman argues that the decline of acoustic recording in the 1910s took place because it 'fell prey to a system-

atic producer campaign to feature continuous musical accompaniment' in preference to recorded sound.<sup>29</sup> Whether this transition was due to a systematic campaign or simply organic market forces, the reason is clear. Recording technology which offered very poor reproduction quality in large auditoria, did not allow editing or mixing and had a maximum playing time of three minutes that clearly could not be used in conjunction with films which by now were frequently over an hour in length and told complex stories through editing and a whole raft of other devices; ones which would require any synchronised soundtrack to be manipulated with an equivalent level of versatility.

The conversion eventually took place when a form of sound technology emerged, and was adapted to the film industry's needs, which did more or less fit that bill. Unlike the acoustic audio used in conjunction with early films, this technology did not come primarily from the record industry; rather, it drew on research carried out by telephone, radio and consumer record industry engineers. What the various components of technology all had in common was the use of electricity to capture and amplify the audio signal, both for storage using an offline carrier (be that a disc, photographically, or later, on magnetic media and as digital data) and for amplification in post-production and playback in the cinema auditorium. It would prove to be ideally suited for overcoming the limitations of acoustic audio.

Electrical audio works by converting the changes in air pressure which the human ear detects as sound into variations of the characteristics of electrical energy over time, which are then recorded. The device which detects those changes in the first place is known as a microphone (or 'mike' for short). The earliest known working example was patented by Alexander Graham Bell – inventor of the telephone – in 1876, and consisted of 'a wire that conducted electrical direct current, with audio signals generated and received via a moving armature transmitter and its associated receiver'.<sup>30</sup> Microphone technology was developed and improved substantially during the following four decades, with increasing sensitivity to changes in air pressure and increases in the signal to noise ratio (the proportion of 'good' recorded sound to distortion in a given recording or transmission). Telephones remained the only application for which microphones were used on any significant scale until the early 1920s, the decade in which radio became established as a mass medium. In the US, for example, regular broadcasts began in 1920 and radio sales 'took off' in the spring of 1922, with over \$60 million in sales of receivers recorded during that year.<sup>31</sup> By the end of the decade, most of the populations of North America and Europe either owned a radio or were able to listen regularly to broadcasts.

By contrast, the recorded music industry stuck entirely with acoustic technology until 1925, and even after electrical recording was introduced, playback in the home remained primarily acoustic until the introduction of microgroove records in 1948. The reason for this was partly cultural and partly technological. Unlike the close ties which exist between the recorded music industry and broadcasting today (for example, music broadcasting and the role of disc jockeys), the two were seen as oppositional in the 1920s. Radio sold itself on being a live medium,

distinct from both the 'canned' media of film and recorded sound. Commercial entertainment broadcasters, therefore, rejected the use of recorded programming as inferior.<sup>32</sup> When the recorded music industry did begin to switch from acoustic recording using horns to electric recording with microphones in the spring of 1925, the impetus came from a sector about as far removed as possible from the popular cultural remit of the now established Hollywood film industry, that of classical music. The improved reproduction quality of electrical recordings was initially exploited in this genre: upon hearing an experimental test record the legendary classical record producer Fred Gaisberg immediately announced his intention to record Wagner operas.<sup>33</sup> As David Morton describes the phenomenon, 'high culture = high fidelity'.<sup>34</sup>

The main technical reason why the manipulation of sound as an electrical signal was restricted to telephones for so long was the problem of amplifying that signal to a level at which it would power a recording device or loudspeaker (i.e. be audibly played back within a large space). Once again, it was the telephone industry which set about trying to solve this problem, in an attempt to extend the limits of its geographical coverage. Most of the main research was carried out during the 1910s: beyond the telephone network, it was commercially exploited by radio broadcasters initially (1920), followed by the recorded music industry (from 1925), and in the cinema last of all (systematically from 1926).

The individual who made the first key discovery and who was instrumental in its subsequent application to film sound was Lee de Forest, an electronics engineer born in Iowa in 1873. On 20 October 1906 he demonstrated a device which he called the 'Audion tube', a vacuum-sealed glass receptacle containing three electrodes, through which a source signal flowed and could be varied (what would now be called a 'triode valve'). Although at the time, he was thinking in terms of using it purely as part of a radio receiver, subsequent research revealed that triode valves could amplify the voltage of an input signal almost five times, while retaining the original pattern of modulation. The resulting electrical current could then be fed to a device which converted the current into audible changes in air pressure – a loudspeaker, in other words – thereby playing back the signal generated by a microphone. The impact of this discovery on all forms of audio technology, not just film sound, was enormous: as Kellogg puts it, de Forest's Audion tube 'unlocked the door to progress and improvement in almost every phase of sound transmission, recording and reproduction'.<sup>35</sup>

De Forest himself was keenly interested in synchronised film sound, and eventually produced the short-lived but technically very successful Phonofilm system (see below). But he could not have adapted his valve amplification technology into a workable system on his own. Instead, the patents to his Audion tube were bought by the American telecommunications giant Western Electric in 1913. Over the following years Western Electric developed amplification technology for use in long-distance telephone transmission, and following the end of World War One further work was done on producing microphones and loudspeakers of sufficient sensitivity and power for use in the then embryonic radio industry.

The chain of events which culminated in the research and development that would lead directly to the mass-rollout of synchronised film sound, together with extensive analyses of the individuals and businesses involved, their roles and their motivations, has been covered exhaustively, painstakingly and authoritatively elsewhere.<sup>36</sup> There would be very little point in my trying to provide a detailed summary of this research here, since this book is principally concerned with the evolution of moving image technology as a whole and with drawing parallels between different branches of it, not with analysing the minutiae of one chapter in that history in order to draw broader conclusions. Rather, we should take the following salient points from that chapter as having a specific bearing on the development of film sound technology thereafter:

- by 1926, four distinct methods – which would subsequently be marketed under the trade names Vitaphone, Movietone, RCA Photophone and Western Electric – had been successfully demonstrated for electrically recording and reproducing audio in synchronisation with a real-time film transport mechanism (i.e. a camera or projector);

- the first significant commercial launch was of the Warner Bros./AT&T/Western Electric 'Vitaphone' system, used in the feature films *Don Juan* and *The Jazz Singer* (1926 and 1927 respectively dir. Alan Crosland). These combined a recorded orchestral score with synchronised effects and, in *The Jazz Singer*, one scene containing a brief exchange of recorded dialogue (which included yet another apocryphal line: 'You ain't seen nothing yet!'). Although the release of these films has been characterised as 'the birth of the talkies' by a number of writers, it is important to note that by this stage, the competing systems were in the final stages of development and almost ready for commercial launch.

- these early experiments were deemed to have made both a cultural and a business case for integrating film with electrical audio recording. In the years 1927–30, the output of film studios in the US and Western Europe gradually moved to including soundtracks. Actuality and other short subjects came first (for example, Fox Movietone News), but by February 1930 95 per cent of Hollywood's output had synchronised sound. The only significant remaining 'silent' areas of production were in documentary, industrial, educational and amateur films.

- the installation of reproduction equipment in cinemas took significantly longer, because of the volume of equipment which needed to be manufactured and shipped, and the financial investment in it was so great (tens of thousands of cinemas, as distinct from ten to twenty studios). In this sector, the conversion process was not totally completed until the end of 1933, and will be covered at greater length in the following chapter.

Within the details of this process, however, can be found some important clues as to the way in which film sound technology would develop during the latter half of the century. Acoustic sound had failed because (i) the continuous playing time was severely limited, (ii) editing was impossible, (iii) synchronisation was unreliable and (iv) it could not be amplified sufficiently for playback in a large auditorium. The form of electrical audio which eventually established itself in the film industry of the 1930s – optical sound-on-film – successfully overcame all these issues.

## The transition process – from experimentation to standardisation

Of the four systems which had been developed in the years leading up to 1926 it was, interestingly, the one which was least compatible with these criteria that was used for Warner Bros.' initial launch. Vitaphone represented a huge step forward from the acoustic technology of the 1900s and 1910s in that it used electrical amplification both to record and reproduce the signal. The signal carrier, however, was still a disc. There were some refinements. The discs were 16 inches in diameter, and rotated at a significantly slower speed – 33rpm compared to the 78rpm which had been universally adopted by the recorded music industry as the standard speed for records for domestic sale in 1915. To compensate for the consequent loss of dynamic range, a thicker and softer shellac compound was used together with a larger groove pitch, which enabled wider and deeper modulations to be inscribed. This produced a comparable signal quality but a longer continuous playing time than the discs' domestic counterparts. However, as a result, Vitaphone discs tended to wear out far more quickly than domestic shellac records. They could be played 'between 18 to 22 times with fairly good results'; according to a projectionists' manual from 1929, cinemas needed to have three sets of discs to last for a full week's run.<sup>37</sup> From Warner Bros.' standpoint, the decision to use disc technology represented a significant time and cost saving, as Western Electric had already invested in the research and development which had led to the introduction of electromagnetic disc cutters – in which the cutting needle is driven by an amplified electrical signal rather than acoustic vibration – in 1925, and these could be adapted for recording synchronous film sound at very little extra cost.

Vitaphone did, however, suffer from some of the same drawbacks as acoustic recording. Each disc surface carried about 10 minutes of playing time (which corresponded to a reel of release print), which all had to be recorded in a single take. Any mixing had to be done 'live' during the recording itself and the disc could not be edited thereafter.<sup>38</sup> As a former radio engineer who moved to Hollywood in the wake of the conversion recalled, 'it was one thing to put a recorded song on a disc, with an ad-lib by Al Jolson at the end of it; it was quite another to edit a fast-moving melodrama, in which there might be a dozen short scenes in one minute of film'.<sup>39</sup> And there were also synchronisation problems. Again, the use of electricity achieved a vast improvement over any of Vitaphone's acoustic predecessors. In the studio, cameras and disc cutters



Fig. 4.1 The projection booth of a cinema in Tooting, South London, circa early 1930s. The projectors are equipped both with optical sound-on-film reproducers and turntables (underneath lamphouse) for Vitaphone records. Courtesy of BFI Stills, Posters and Designs.

were driven by self-synchronous (selsyn) motors, in which the frequency of the AC source power supply ensured that their speed remained identical. In the cinema, 'the turntable carrying the disc is on the same motor shaft that drives the mechanism of the projector',<sup>40</sup> theoretically ensuring perfect synchronisation. But though it was much improved, this arrangement was not foolproof (especially at the exhibition end, which will be covered in the next chapter).

In order for synchronised sound to be fully compatible with the production techniques which had been evolved by Hollywood during the 1920s, a system was needed which, like Vitaphone, exploited the possibilities of electrical recording and amplification; but which, unlike Vitaphone, did not suffer from the drawbacks of using discs as the carrier (unreliable synchronisation, limited running time and inflexibility in editing). This technology was already in an advanced stage of development by the time Vitaphone was launched to the public, and would push it out of the marketplace altogether by the early 1930s. This was 'optical' sound, a technique which worked by creating and reproducing a photographic record of the electrical impulses produced by a microphone.

The technique of representing soundwaves photographically (yet again) was not new. Possibly the first person to have discussed the idea in print was Professor E. W. Blake of Brown University. In 1878 he produced 'photographic records of speech sound on a moving photographic plate, using a vibrating mirror',<sup>41</sup> though there is no evidence that he managed (or even attempted) to reproduce them as audible sound. As with electrical disc recording, it was the discovery of Lee de Forest's Audion tube and the subsequent development of microphone technology which really opened up the possibility of using photographic film to record sound as well as pictures. In crude terms, this technique works as follows: as with telephone or radio transmission, and disc recording, sound is captured by a microphone and the resulting electrical impulses are amplified by electrically, by means of valves. However, instead of feeding this signal to a disc cutter to produce a groove in a record, it is used to control the flow of light to a strip of unexposed film, which passes in front of the light source at a constant speed. The stronger the signal, the brighter the light, and vice-versa. When processed, the exposed film therefore carries a permanent record of the modulated signal over time, one which is theoretically capable of being reproduced as audible sound.

In practice, working systems for both recording and reproduction did not become available until the early 1920s. This was because, even though it would be theoretically possible to regulate a light aperture controlling exposure acoustically, optical sound was wholly dependant on electrical amplification for playback. As there were no indentations in the processed sound film, the only way it could be played back was to find a means of converting the processed photographic sound record back into an electrical impulse for amplification. A key discovery toward making this possible was demonstrated by the British engineer Willoughby Smith in 1873, which was that the metal selenium generated small amounts of electrical energy in proportion to the intensity of light it was exposed to. This eventually led to the production of the 'photoelectric cell', which forms the basis of the optical sound reproduction

head in all cinema projectors manufactured from the early 1930s onwards. However, it was impossible to exploit this discovery in the absence of any means of electrical amplification for playback. Though de Forest's work cleared the way for optical sound to become a reality, systematic research did not get underway until the late 1910s, primarily because the telecommunications industry initially had other research and development priorities for the Audion tube, and thereafter World War One put a stop to most technical research and development in the US and European film industries.

By the start of the 1920s, a number of inventors were working on different systems, backed by a combination of industrial capital and private investment. As we have seen above, the Warner Bros./AT&T Vitaphone system was the first to bring electrical recording and reproduction to the market, primarily because it represented a cleverly integrated combination of two 'off the shelf' technologies initially developed by separate industries – the amplification technology found in telecommunications and radio and the audio carrier used by the retail music industry – rather than a fundamentally new technology which was specifically tailored to the needs of film studios and cinemas. This gave it a slight head start. There were also no less than three optical systems being developed simultaneously, the most successful of which (albeit in a much modified form), remains in widespread use well into the twenty-first century. These were, briefly:

- a system which resulted from work carried out by Lee de Forest and another American, Theodore W. Case. In 1917 Case had discovered a significantly more sensitive photoelectric compound than selenium, and by 1922 had developed the 'AEO-light', a hydrogen-filled bulb which responded to variations in electrical current. Combined with de Forest's amplification technology, the package was christened Phonofilm and was first publicly demonstrated in April 1923 in New York. In the Phonofilm sound camera, the intensity of exposure through a fixed aperture varied according to the current applied to the light. As the developed sound record was thus of fixed width but with varying modulation according to the density of silver salts remaining in the emulsion, it was known as a *variable density* soundtrack. For the next few years Phonofilm was used on a small scale to make a number of musical and comedy shorts ('about 34 theatres' in the US-installed playback equipment, according to Coe,<sup>42</sup> and the system was also used on a limited scale in Britain), until the Fox studio eventually bought the rights in July 1926. By the following year Fox was starting to release full-length Movietone features on a regular basis, and from autumn 1927 a bi-weekly newsreel with synchronised sound (something which the logistics of Vitaphone recording would have made impossible).

- a system which recorded optical sound by varying the size of the aperture rather than the intensity of the light source. This produced an exposure of equal intensity, but with variations in the signal modulation being registered as movement of the boundary between the exposed and opaque areas of the recording. This,

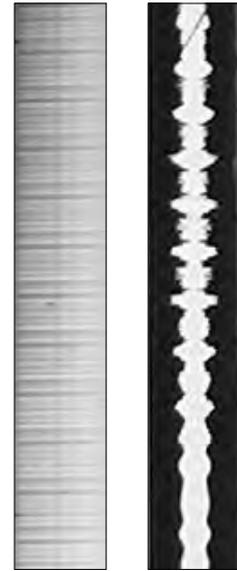


Fig. 4.2 Variable density (left) and bilateral variable area (right) optical sound records.

therefore, was known as *variable area* sound. The invention which facilitated it was a modified galvanometer which caused a small mirror to vibrate in response to the input signal. When a uniform light source was projected onto this mirror, it was reflected through an aperture and onto the emulsion of the passing unexposed film stock in proportion to the source modulation. This technique was first demonstrated by the scientist Charles Hoxie in 1921 and subsequently refined by researchers working for the Radio Corporation of America (RCA), which bought the rights. At the time RCA was effectively a front organisation for two large American industrial combines, Westinghouse and the General Electric Corporation, both in exploiting new audio technologies and also as one of the emerging broadcasting majors. The problem RCA had with Photophone was that between them, Warner Bros. and Fox had effectively saturated the (then nascent) film sound market. RCA's solution was to start its own studio and cinema infrastructure, formed by buying out a large theatre chain (Keith-Albee-Orpheum, hence 'KO') and with RCA providing the sound technology. RKO, therefore, is unique among the Hollywood majors of the 1930s and 1940s in that it was established with the sole aim of commercially exploiting sound technology. Initially only RKO used Photophone among the Hollywood majors, although the system quickly gained a growing market share among producers of B-pictures, newsreels, documentaries and animation. Ironically, the second major studio to adopt it for feature film production was Warner Bros. itself, having finally abandoned Vitaphone in the early 1930s and variable density recording in 1936.

- a refined variable density system developed by Western Electric as it became clear that the technical advantages of sound-on-film would push disc recording out of the market. In this method the modulation was not recorded by varying the brightness of the lamp making the exposure, but by means of a 'light valve', which varied the width of the slit aperture which regulated it. As it was lighter than the RCA mirror and worked at a much lower voltage than the AEO light, it was far more sensitive to changes in modulation – especially at lower volume – than either of the other two optical systems. In 1928 the five major Hollywood studios agreed to adopt this new standard for subsequent feature production.

Of the four systems which had been successfully demonstrated by 1926, the years between 1928 and 1932 saw two (Vitaphone and Movietone) fall by the wayside and the other two (Western Electric and RCA) establish market shares which they would maintain and develop for two decades subsequently. Vitaphone discs remained in use as a playback medium in cinemas for 2–3 years after the system had been abandoned as a production medium, largely because of the investment which cinemas had already made in the equipment (this issue will be covered in greater detail in the next chapter). Compared to the Western Electric method, the sound quality recorded by the AEO light was 'poor, and the Fox-Case system soon went the way of Vitaphone and the carrier pigeon'.<sup>43</sup> Of the two technologies which remained, Western Electric dominated Hollywood throughout the 1930s. RCA, however, established far more of a foothold in Western Europe (especially in the UK and France, where variable area recording, either using RCA equipment itself or sub-licensed forms of it, soon dominated those studios) and South America. It soon be-

came apparent that variable area recording offered significant technical advantages over variable density, of which more below.

### The 'key moment'

The historian Thomas M. Cripps notes that:

The chronicle of soundfilm was no more risk free than the stories of other technologies. But the manner of its success differed from almost all other cinematic achievements. Neither the adventurers in the banks nor the visionaries in the lab could have predicted soundfilm's tremendous success after a vaudeville-like snippet of it premiered in the summer of 1926.<sup>44</sup>

He is certainly correct to identify the speed and universality of the conversion process as being unique among the complex and interrelated instances of technological, economic, industrial and political forces determining the when, where and how of technological change in moving images. Here are some examples by way of comparison: camera and projection speeds took nearly three decades to standardise. Photographically recorded and reproduced colour took almost four decades to make the jump from a prestige, niche-market technology to being almost universal in film and television production. The 'Academy' aspect ratio was abandoned as a universal standard for film in the early 1950s, and over half a century later, no single wide-screen standard has emerged to replace it. In almost eighty years of regular broadcasting there has never been a single, worldwide standard for the signal format of an analogue television transmission. Compared to these timescales, the practice of producing a synchronised and mixed soundtrack supplied alongside the moving image, and recorded in a way that was universally replayable on a wide range of cinema equipment, took place almost overnight. The only other conversion processes to have happened anything like this quickly were the introduction of panchromatic film and the nitrate to acetate conversion (see chapter one). There are some notable similarities and differences between the two, of which more below.

As stated in the introduction, the 'key moment' has been heavily researched, discussed and mythologised; probably more so than any other process of technological change in the history of the film industry. In fact, it is probably the *only* process of technological change which the lay cinemagoer is likely to be aware of, in any meaningful historical sense. Support for this contention can be found in the fact that of ten people I recently spoke to at random in the bar of my local arthouse cinema, eight had not even heard of the term 'CinemaScope' (and the other two were not able to define it to any degree of accuracy), while all but one knew that sound had been 'invented' in the late 1920s. Seven respondents gave their source of information as the film *Singin' in the Rain*. Even at the time that film was set, one of the proliferation of technical manuals for projectionists published in the wake of the conversion noted the extent of the public reaction it had generated, commenting with frustration that 'every now and again the newspapers and the trade publications hail some individual

as having been the first to have invented “sound pictures”<sup>45</sup>. So why is this the case, and do the unique aspects of the ‘key moment’ set it totally apart from other instances of technological change in the history of moving image technology?

The first unique factor to note is that the ‘key moment’ was highly visible and heavily publicised. Hollywood had spent the previous decade developing the ‘classical’ continuity system of shooting and editing, the main aim of which was to make the role of technology invisible. As a member of the cinema audience you were not supposed to know or care whether a scene had been shot using orthochromatic or panchromatic stock, how it had been lit or even if it had been shot on location or a studio set. Sound, on the other hand, was a rare example of technology – and specifically the idea of technological change – being heavily promoted and marketed to consumers of the film industry’s output. True, this also happened to a limited extent with Technicolor and early widescreen as well. But the publicity associated with these technologies was aimed at identifying the films and cinemas which offered them as prestige, niche market products, as exceptions which proved a rule. This was very similar way to the phenomenon David Morton identifies, of marketing of hi-fi classical music recordings by the retail music industry when electrical recording was introduced in the mid-1920s and then long-playing records in the early 1950s. In contrast, the promotion of movie sound was explicitly pitched as a new technology which was here to stay and which would engender permanent change.

To a certain extent, accounts of the conversion such as the one found in *Singin’ in the Rain* represent the end result of a legend becoming fact and then being printed as such. It will already have become apparent by now that most published research on the ‘key moment’ – including this summary of it – deals almost exclusively with events that took place in the United States. A demonstration of the extent to which the Hollywood ‘key moment’ myth has come to dominate can be seen by looking further afield, where the conversion process starts to extend and become more complicated. In Western Europe, for example, patent and licensing disputes inhibited both the conversion process itself and the evolution of film sound technology thereafter. Eventually, compatibility issues and licensing restrictions associated with the German ‘Tri Ergon’ variable density system were eventually used as a political weapon by the Nazis, as a way of restricting the import of Hollywood films.<sup>46</sup> In the Asian subcontinent, where the American giants could not enforce their patent rights as effectively as they could in the developed world, the conversion process happened a lot more slowly, as indigenous companies developed their own equivalents of the Hollywood technologies: in Japan, for example, 14 per cent of feature films were still being shot silent by 1942.<sup>47</sup>

A second reason for the ‘key moment’ being understood as a unique phenomenon was its sheer cost in relation to the speed at which it took place. As argued in chapter two, a combination of the depression and the time taken to amortise the investment in sound almost certainly inhibited the commercial development of widescreen for two decades subsequently, and may well have done likewise both with colour (see chapter three) and stereo (see below). This investment was not so much in the research and development of the technologies themselves: in this

regard the film industry was following one of a long line of precedents of cashing in on technical innovation which had been generated either by small scale 'cottage industry' research by private individuals, or by adapting technology produced by big business in other related industrial sectors, or – as in this case – by a combination of the two. The big bucks were spent on installing equipment in cinemas. I shall discuss this process at greater length in the next chapter; it will suffice to note at present that by 1929, there were four systems on the market being used by Hollywood to record sound, and over 300 for reproducing it in the cinema. The asset base of Warner Bros., for instance, grew from just over \$5 million in 1925 to \$230 million by 1930.<sup>48</sup> It was a similar story with the other vertically integrated majors, not to mention the thousands of independent exhibitors which found themselves having to buy equipment in order to stay in business. Most of this infrastructure was paid for by speculative investment, and it all happened at the height of one of the biggest economic booms in America's history, when such investment was easy to come by. As Douglas Gomery notes, 'Warner Bros. 'had the support of America's most important banks throughout this period of expansion. And sound was only one part, albeit a very risky one, of the investment surge [of the mid-1920s]'.<sup>49</sup> Despite the overlap in technology with other allied industries, this electronic hardware and the manpower installing it could not be removed and put to other use. Once the conversion bandwagon had gathered a critical mass of momentum, there was simply no going back (apart from writing off a colossal investment) even if the industry and/or its bankers had wanted to do so, especially after the Wall Street Crash. By contrast, the technical perfection of widescreen and stereo did not happen until shortly after the bust which followed. In economic terms, the commercialisation of these two technologies was set back to where sound had been two decades earlier, even if, by the early 1930s, the technologies themselves were considerably more advanced.

However, there are some other aspects of the 'key moment' which do have similarities with other examples of technological change, most notably the nitrate to safety conversion. The first – and this flies in the face of a lot of 'key moment' mythology – is that it was achieved largely on the back of off-the-shelf technology. With the sole exception of the hardware for recording and reproducing optical sound (and that was only developed to the point of marketability as a direct response to the obvious shortcomings of the alternative), virtually all the technology used to make the 'key moment' happen had been developed for use in telecommunications, broadcasting and/or the retail music industry. It was adapted for use with films when it became clear that it would enable the restrictions of acoustic disc-based recording to be overcome and thereby allow recorded sound to meet the requirements of the 'classical' production conventions being evolved by the mainstream studios. Safety film had followed a similar pattern. Attempts to inhibit the flammability of nitrate went back at least to 1904, but for over four decades subsequently it made more economic and technological sense to manage and minimise the risk of fire than to adopt an alternative which was more expensive, less reliable and to all intents and purposes just did not do the job. With both sound and safety film, there were important points along the way: de Forest's invention of the triode valve, improvements

in microphone technology, the mass-manufacture of diacetate and butyrate for use in small-gauge stocks, or the introduction of propionate in 1938. It is less clear with safety film than with sound as to what precisely tipped the scales; but as with sound the conversion happened astonishingly quickly once that point was finally reached. There are other examples of similar chains of events, but none which took place in anything close to the time scale as that of the 'key moment' or the introduction of safety film.

The second aspect of the 'key moment' which bears similarities to the mass-rollout of other moving image technologies is in the role of standardisation, which enabled sound to integrate with dominant business models. Barry Salt expresses the rationale for Bell's 'fondest hope' in somewhat more pragmatic terms:

The almost inevitable sequence of events is that one company chooses a standard for its own use, and then, either because the company concerned was first in the field, or because it establishes economic superiority, or because the standard is obviously a sensible and practical one, the other companies adopt it as well. Only after this has happened is the standard ratified by an industry body.<sup>50</sup>

In other words, market forces do the job, the end result of which is then enshrined by formal technical standards, which in turn derive their authority from an already-established mass acceptance. In the case of sound the pioneer systems – Vitaphone and Fox/Case – ultimately failed, for exactly the same reasons that (for example) Technicolor and the CDS sound system system (see chapter 5) eventually followed suit: they proved the business case for the deliverable, but then someone else came along with a more efficient way of delivering it. Initially, the four sound systems were totally incompatible with each other at the exhibition end. A film recorded using Vitaphone equipment could not be shown in a cinema which used Western Electric reproducers, and vice-versa. The studios very quickly realised that if this situation had been allowed to continue, it would push costs up and reduce revenues unnecessarily. Cinemas which could afford them would install multiple sound systems, which would benefit the electronics manufacturing industry at the expense of the Hollywood studios, but other cinemas would not be able to show films which they otherwise might have done, thereby reducing the earning power of the studios' product. As with 35mm film itself, it was for these reasons that the five major studios agreed in 1928 to standardise certain technical characteristics of optical sound implementation between the rival products available to studios, principally the film transport speed (24fps), the position and width of the soundtrack and the 'offset', i.e. the distance between the synchronisation points for picture and sound (for 35mm film the sound was offset 20 frames ahead of the picture). This ensured that, while there may be differences between the cost and quality of the various systems available, any soundtrack which adhered to the standards would be playable in any cinema throughout the world. That is not to say that compatibility issues never surfaced thereafter. In particular, there is evidence to suggest that the uptake of specific sound systems was used by the Hollywood studios to gain economic advantage. In

the early 1930s, for example, allegations surfaced to the effect that British studios were being refused access to American-owned distributors unless they used sound recording equipment imported from the US.<sup>51</sup>

By the early 1930s, therefore, moving images and electrical audio recording had become two inextricably linked technologies. An interesting indication of just how quickly the 'key moment' had taken hold can be found in the existence of some voices (though admittedly, not many) who still believed that this link was not necessarily inextricable, that silent cinema had not gone forever and that the two could look forward to a long period of co-existence. One example was the British critic Ernest Betts, who, writing in the highbrow journal *Close Up*, opined that sound would become the established norm for newsreels, actualities and other forms of short film, 'because of their novelty and magnetism of the human voice, such items will always have a place in our film programmes',<sup>52</sup> but that synchronised sound was unsuitable for features, the majority of which would continue to be silent. But articles such as his were increasingly becoming exceptions which proved the rule.

### Consolidation in the 1930s and 1940s

The following two decades present a similar pattern to that of the previous three: they consist primarily of the consolidation of existing technologies into mainstream industrial practice, and the beginnings of the research and development process which would eventually bring new ones to the marketplace.

As has been mentioned above, the early part of the 1930s would see Vitaphone and Movietone disappear and Western Electric variable density and RCA variable area sound (and derivatives thereof) become the principal recording technologies in use worldwide. Although Vitaphone had effectively been abandoned as a studio production medium during 1929 in favour of Western Electric, disc versions of the final soundtracks continued to be released to cinemas until 1931–32, when they were phased out (more on this in the next chapter). With the exception of RKO, the other studios also adopted Western Electric. RCA Photophone quickly established a dominating market position in Britain, while the rest of Europe saw the rapid growth of a German variable density system known as Tri-Ergon, backed by a well-capitalised consortium of European investors, the Tonbild-Syndikat (Tobis). As a result, Europe witnessed what was almost a full-scale trade war with the Americans over sound systems. It was eventually resolved at a conference held in Paris on 22 July 1930 between key US and European representatives, who agreed to operate a worldwide cartel for the export markets of sound technology. Most of mainland Europe became the exclusive territory of Tobis; though once again, a key provision of the agreement was that of universal compatibility between playback systems. The British market was unrestricted, with the result that a number of home-grown systems emerged. Notable among them were British Acoustic Film, launched in 1925 and based on variable area technology developed by the Danish engineers Axel Petersen and Arnold Poulsen, and Visatone-Marconi, which was in use by the mid-1930s. Both were used extensively in newsreel, documentary and low-budget feature production. With

minor modifications, the Paris agreement continued to operate until the outbreak of World War Two.<sup>53</sup>

In the Hollywood studios, gradual refinement of the Western Electric and RCA systems continued. It was noted earlier that one of the main reasons for Vitaphone's demise was inflexibility in editing. The first generation of Movietone and Photophone recording equipment was not much better: it was 'single system', meaning that both the image and the optical sound record were exposed simultaneously, in the same camera and onto the same strip of film. The advantages were foolproof synchronisation and a reduction in the amount of equipment needed for location work (hence Movietone's initial launch in newsreels), but the possibilities for creative editing were still limited. As editing optical sound film was accomplished in exactly the same way as editing the picture – i.e. by cutting and joining it – any cut made to the original negative would affect both the picture and the sound. Add to that the 20-frame offset and it quickly became clear that for everything except footage that would only need minimal sound editing (for example, actuality footage), single system was simply too restrictive. Possibly the most widely known single system feature film (and one of the most heavily analysed early sound films in existence) is *Blackmail* (1929, dir. Alfred Hitchcock). The fact that the single-system Photophone camera made post-synching impossible was exacerbated by the fact that the lead actress played the part of a London shopkeeper's daughter, but was a Czech who spoke no English! The production had started life as a silent in which this obviously was not a problem. When the studio took delivery of some of the first RCA equipment to reach Britain midway through shooting, a last-minute decision was made to release both the film as planned and a talkie version, the latter containing a number of scenes specially reshot for sound.<sup>54</sup> As the use of single system made it virtually impossible to post-dub an English-speaking actress (to achieve this, all the other components of the soundtrack would have had to be recreated in real time along with her voice), she spoke the lines into a microphone off-screen while they were mimed by the Czech actress on camera. As a result of logistical problems such as these, one of the many film editing manuals to have dissected *Blackmail* in minute detail observes that 'the difference between the two versions is palpable. The freedom available to shoot and edit without constraint is manifest in the variety of structuring possibilities that Hitchcock exploits in the silent version.'<sup>55</sup>

Within a few years single system had been superseded by the use of separate 'sound cameras' (devices which expose an optical sound record onto negative stock independently of the camera used to shoot the picture), with clapperboards providing a reference point for synchronisation (see chapter two). The separate sound negatives could be cut and joined far more flexibly and in perfect synchronisation with the picture, as there was no fixed offset in the recording. Furthermore the opportunities for mixing were to all intents and purposes unlimited, as two or more reproducers (each running film which had been edited as necessary) could be adjusted to differing volume levels, with the combined output from all of them fed to a sound camera which exposed a new negative carrying the mixed recording. While this did introduce the generational fading which is common to all duplication

of analogue signals, various methods were evolved for keeping the loss of signal to noise ratio (i.e. sound quality) to a minimum. Most notable among these were that film manufacturers developed stocks which were specifically designed for optical sound mastering and duplication. While the introduction of panchromatic film in the late 1920s (see chapter one) greatly expanded the creative opportunities available to cinematographers, it was not the ideal type of film emulsion for optical sound recording. As the exposure in a sound camera came from an artificial light source, a more accurate sound record could be obtained by using film stock which was photosensitive only to those areas of the colour spectrum in which the sound camera generated light. This obviously was not an option with single system, because the same strip of film had to carry both the picture and the sound. By 1932, both Eastman Kodak and Du Pont (the two major American stock manufacturers) were marketing fine-grain stocks optimised specifically for variable area and variable density sound mastering.<sup>56</sup> The sound negatives these produced could be copied optically (i.e. by contact printing) with negligible loss of signal, or electronically if mixing was needed.

Of the two methods of optical sound recording which became established during the 1930s, variable area gradually emerged as the technically superior system. Initially the RCA sound cameras had suffered from what was termed 'volume expansion' by the sound engineers who worked with it. The mirror which directed the light beam in response to the input signal was significantly heavier than the equivalent light regulation device (the 'Carolus cell') in the Western Electric variable density sound camera. It failed to respond effectively to low signal levels (meaning that quieter sounds were less audible), but over-reacted to higher ones, resulting in overmodulation. This was gradually overcome by progressive refinements to the RCA galvanometer system throughout the 1930s and early 1940s, eventually leading to the use of ultra-violet light to make the exposure in the RCA sound cameras marketed in the late 1930s. Its much lower wavelength than that of light within the visible spectrum vastly increased the strength of the recorded signal and its durability through successive generations of copying.<sup>57</sup>

The main flaw with variable density was not so easily overcome. It depended on variations in the chemical composition of the emulsion across the topography of the film to accurately represent the sound record. Variable area, however, only needed the film to register the difference between opaque and transparent – it was the accuracy in where that boundary lay which determined the quality of reproduction. Therefore, a density track was far more likely to be significantly degraded through successive generations of copying than an area one, because quality control in the lab was far more critical. As one archivist who has spent a significant proportion of his career working with variable density soundtracks (and who describes himself as 'possibly the last person on this planet to re-record using a variable density sound camera') stresses: 'control, or, if you wish, grading of variable density sound is both vital to both quality and volume of duplicated sound materials. Signal to noise and distortion can become enormous problems if the closest of attention is not paid to sound control.'<sup>58</sup> Furthermore, a number of noise reduction techniques for variable



Fig. 4.3 Schematic of an early RCA variable area recording mechanism.<sup>59</sup>

area systems were introduced throughout the 1930s which further enabled almost lossless copying (thereby increasing the creative possibilities for editing and mixing), notably the 'push-pull' soundtrack, while their variable density equivalents proved to be far less effective. However, due to patent restrictions and contractual arrangements between studios and the owners of sound technology, Western Electric maintained its dominating market share as Hollywood's principal film sound system until the advent of magnetic recording in the late 1940s. Optical sound would, however, remain the primary method of reproducing film sound in the cinema up until the time of writing. The main developments in this technology in the post-war period were concentrated almost exclusively in the area of release formats for cinema exhibition, and it is for this reason that they will be discussed in the next chapter. For the purposes of this one, it is enough to note that as the technical superiority of variable area became apparent, the RCA system and derivatives thereof gradually increased their market share. The use of variable density declined significantly in the late 1940s, and had all but disappeared from mainstream use by the end of the 1950s.<sup>60</sup>

### Magnetic sound in the 1950s

The use of magnetic tape to record and edit film sound was the next major technological development to significantly affect the production and exhibition of film and television. Optical systems captured a record of the modulating input signal using photographic emulsion. Magnetic sound does so on a metallic compound which is sensitive to magnetic fields (initially iron oxide, but subsequent refinements introduced magnetic media of greater sensitivity, such as chromium dioxide) coated, like film emulsion, on a flexible base. The input signal is fed to an electromagnetic 'head' while the coated base passes in contact with it at a constant speed. The head converts the input signal into magnetic energy, which in turn causes the oxide particles to change their relative positions (i.e. some particles are positively charged

while others are not) in a way that creates a record of the input signal. Playback is simply the reverse of this process: the coated tape or film is passed across the head a second time, only without any signal being fed to it. This time the head responds to the changing patterns of magnetised oxide on the tape, and the resulting signal is electronically amplified in order to play back the recording.

The development and commercial rollout of magnetic sound followed a similar pattern to that of the other technologies discussed in this chapter. Early attempts by engineers to put theory into practice stretch back well into the nineteenth century. The first one which technological historians generally cite as having been successful was a magnetic recorder demonstrated by the Danish engineer Valdemar Poulsen at around the turn of the twentieth century. His 'Telegraphone' used steel wire as the recording medium. It was of limited use because the absence of any means of amplification meant that the quality of the recorded signal was low and that it could not be played back at any significant volume. There is no evidence to suggest that magnetic recording technology underwent any major developments between then and the 1930s, though a number of scientists and engineers worked on it experimentally.

Once the conversion to sound was complete, the film industry realised early on that this method offered several potential advantages over optical sound. Although the experimental magnetic recorders built before the 1940s generally produced much poorer sound quality than their state-of-the-art optical equivalents, research in the post-war period quickly enabled magnetic sound to record a much wider frequency range and with much lower noise than any optical system. And just as optical sound had provided a number of significant advantages over acoustic and electromechanical disc recording, so did magnetic over optical. Although optical sound could be cut, edited and mixed without practical limit, the process of doing so was costly and time consuming. This was because, like discs, optical sound was also a 'write once' medium: the film stock could only be exposed once, and the sound record could only be played back after it had been sent to a lab for processing. Although a defect in an optical recording could be corrected by replacing only the affected element or section (unlike with discs or single system, in which all elements of the affected reel or shots – including the picture – had to be retaken), playback or mixing was not instantaneous. With magnetic sound it was: as with a modern tape recorder, all the sound engineer had to do was to rewind the reel and press the play button. Furthermore the tapes could be erased and re-recorded ad infinitum. Noting these points, one American writer suggested in 1931 that:

What seems to be the most practicable method thus far suggested is to record the electrical variations as changes in magnetism. If it could be perfected, some apparatus such as the electromagnetic recorder, known as the Telegraphone, invented years ago by the scientist Poulsen, could be used.<sup>61</sup>

In fact, the next major developments in magnetic sound appear to originate (as with safety film and tripack colour) with the Nazis, who seem to have been more in-

terested in its potential for use in radio broadcasting than as a means of adding sound to moving images: as Winston argues, 'it was not until after World War Two that American interest in magnetic recording was revived with the importation of Nazi "Magnetophon" recorders'.<sup>62</sup> In contrast to the reluctance of American broadcasters to use recorded sound because they promoted radio as being a primarily 'live' medium, the Nazis found it ideally suited to many of their propaganda requirements. One interesting footnote to this story is that the Allied intelligence services were persistently unable, throughout the war, to identify the geographical origin of many 'black' propaganda broadcasts to the UK: they were unaware of the advances which had been made in magnetic sound and therefore assumed the broadcasts to have been live because their quality was too high for the source to have been a disc recording. While the early models of recorder (such as the 'Blattnerphone' of 1929, which was also used on a limited scale by the BBC in the 1930s) used by Nazi broadcasters also recorded onto steel wire, the development of a flexible base was quickly identified as a priority, mainly because of the relatively low sensitivity of 'raw' steel, but also because of the health and safety risks of using sharp steel wire in a high-speed transport mechanism. In 1934 the German chemical giant Badische Anilin und Soda Fabrik (BASF) started to manufacture tape on a cellulose diacetate base (with the oxide being supplied by another infamous name in Nazi heavy industry, I. G. Farben), and in 1938 German state radio began using modified versions of a later Poulsen-designed recorder, labelled the Magnetophon.

As with the Agfa tripack colour system, this technology was appropriated by the Americans in the immediate aftermath of the war. The Ampex corporation (of which more in chapter six) quickly established itself as the US's key manufacturer of equipment, while the Minnesota Mining and Manufacturing Company (later 3M) was the first industrial-scale producer of tape in the West, initially on a triacetate base but subsequently on polyester. Even as late as the early 1950s, the embryonic Sony company in Tokyo was still experimenting with paper as the base,<sup>63</sup> while in Britain magnetic recording was not established among broadcasters and the music industry until the mid-1950s.<sup>64</sup>

The film industry moved quickly. By 1948 both Western Electric and RCA were marketing magnetic recorders which were designed to integrate with the existing production practices of studios that already used their optical equipment. The initial recording, editing and mixing took place on magnetic stock. At the end of the whole process, an optical sound negative was exposed of the final mix, which was used to add a combined optical soundtrack to the release prints. The efficiency of this process was further enhanced when the studio engineer Norman Leevers came up with the idea of coating conventional 35mm raw stock with magnetic oxide and making the recording at the same speed as photographic film passed through a camera (i.e. 1½ feet per second, the equivalent of 24 frames assuming a four-perf pulldown), meaning that the sound record could be run, cut and joined in synchronisation with the picture film during editing, just as optical soundtracks had.<sup>65</sup> The disruption to studio techniques and the retraining needed by recordists and engineers was therefore minimal.

By the late 1950s magnetic recording had become standard practice in all forms of production, from Hollywood blockbusters to home movies, and would remain so until the film and television production industries turned to digital sound recording in the mid- to late 1980s. Soon, ¼-inch or ½-inch tape recorded on portable machines quickly became the norm for location recording, while an updated form of single system was adopted for television news and documentary shooting: the ‘combined magnetic’ track, in which a ‘stripe’ of oxide was applied to unexposed 16mm film stock in the space normally occupied by the optical track. Since the content of this track could be copied onto another magnetic tape or film for editing after processing and then the final mix rerecorded to the original element after it had been cut if necessary, the magnetic single system was a lot more flexible than its optical predecessor.

The rollout of magnetic sound as a playback medium for cinemas followed a significantly different pattern. In fact, the introduction of this technology broke a long-standing commercial link between the manufacture and supply of sound-recording equipment for studio production and for cinema reproduction. The two were no longer packaged in the same way as had been the case during the period when optical recording was in mainstream use, and the cinema exhibition industry proved increasingly reluctant to invest in buildings and equipment upgrade. Where new technologies were rolled out, backwards compatibility became an absolute prerequisite, as will be discussed further in the next chapter.

### **Stereo sound in the film industry: 1952 to the present day**

Multi-channel or ‘stereo’ sound for films was also nothing new by the time of its first commercial rollout in the early 1950s.<sup>66</sup> Examples of early research into the idea date back to October 1911,<sup>67</sup> though it is generally agreed that the first successful demonstrations took place during the 1930s. In Britain, the telecommunications and radar engineer Alan Blumlein produced a sound camera which contained a ‘totally new’ galvanometer. It enabled two variable area optical tracks to be recorded in the space previously occupied by one, and Blumlein carried out extensive tests with various techniques for positioning microphones.<sup>68</sup> The subject matter of his ‘binaural’ films was clearly designed to reproduce a sense of spatial positioning when played back through loudspeakers placed to the left and right of the auditorium. Of the surviving films, the two most effective (both shot in July 1935) show steam trains passing through a suburban station in West London, and a horse-drawn fire engine travelling across a field, the bell on which can clearly be heard panning from left to right. Similar experiments were carried out by Bell Telephone Laboratories in the US, supervised by the engineer Harvey Fletcher, eventually leading to what is generally believed to have been the first mainstream feature film to be recorded and presented in stereo. *Fantasia* (1940, dir. Ben Sharpsteen *et al.*) was a feature-length cartoon, the soundtrack of which consisted mainly of orchestral music. The music was recorded using eight channels and then mixed down to four for exhibition, with the stereo mix being played in a small number of ‘road show’ venues which had installed the necessary equipment.

This appears to have set a precedent for stereo in more ways than one. The 'high culture = high fidelity' phenomenon identified by David Morton in relation to the introduction of electrical recording in the 1920s repeated itself with stereo, both in the consumer audio and film sound markets. In the former, the 'long-playing' microgroove record (LP), introduced in 1948, was 'envisioned with classical music in mind'<sup>69</sup> while stereo recordings, first sold as ¼-inch tapes from 1955 and, following the launch of the Western Electric stereo disc cutter in 1958, as LPs, were marketed aggressively at middle-class consumers with high disposable incomes. Once again, classical music was mainly responsible for the sector's growth, in the form of a 25 per cent surge in LP sales between 1959 and 1961. Unlike the 1920s 'key moment', when electrical recording had been redefined from an elitist consumer technology to a mass medium in the form of film sound, stereo sound in films would remain a high value niche market in films as well, until the introduction of Dolby stereo variable area optical tracks in 1976 (which is covered in the next chapter) shifted the economies of scale to enable a mass-rollout of stereo in the cinema.

The systematic introduction of stereo recording and mixing within the film industry took place in the early 1950s. While the techniques and technologies for studio production practices were gradually evolved and standardised, two key factors prevented any form of it from becoming a universal standard. Firstly, all the delivery systems for playback in the cinema were magnetic, and therefore required significant additional investment at the exhibition end. Secondly, all the early stereo systems were packaged with widescreen. Cinerama, CinemaScope and Todd-AO all included their own reproduction formats, which required dedicated equipment and which were incompatible with any of their rivals. While the 'Perspecta' pseudo-stereo optical system, which extracted directional sound information from a single channel, was used to a limited extent in conjunction with the early widescreen systems (most extensively with VistaVision), stereo remained an exception which proved the rule until the latter quarter of the decade. Following the lead established by *Fantasia* and the popular perception which was established in the 1950s of stereo being first and foremost a medium for high fidelity music reproduction, most of the 'road show' prestige feature films released during this decade and featuring stereo sound (many of which were musicals anyway) used the stereo spread predominately for music, and in some cases only for music. As John Belton notes, 'critics, industry personnel and audiences accepted stereo scoring, while rejecting directionality for dialogue'.<sup>70</sup> Over the past five decades the use of a centre channel, specifically for dialogue and nothing else, has gradually become an almost universal convention, both in studio mixing and cinema reproduction. Its use became widespread following the introduction of Dolby SVA in 1976, which standardised four channels (left, centre, right and surround) as the industry norm for the subsequent two decades. Rick Altman argues that:

Listening to the centre channel is like listening to a telephone during a music concert, simultaneously satisfying our expectations for music reproduction (large room with high levels of long, slow reverberation and a wide frequency range) along with the standards that we

have learned to apply to dialogue transmission (spacelessness and no reverb, with a relatively narrow frequency range).<sup>71</sup>

With stereo sound, therefore, we have witnessed the same process of evolution which characterised the 'key moment' and then the move to magnetic: from a form and use of technology which was either experimental or borrowed from other industries (in this case, the retail music industry) to one which addressed the production and reception context of film specifically.

Developments in film sound technology from the mid-1970s to the turn of the century are concentrated mainly in the area of cinema exhibition, and so will be covered in the next chapter. In the context of production, the 1970s and 1980s saw stereo become the norm as cheap, reliable and backwards compatible equipment for cinema playback became available to support its use in shooting and post-production. The number of channels and the dynamic range possible from magnetic recording gradually increased, and a number of new developments further cemented the use of analogue magnetic recording throughout the 1970s and 1980s. A number of very high quality and easily portable ¼-inch magnetic tape recorders were developed, mainly in Europe, including the Swiss Nagra and the German Uher. Although these machines were initially used mainly in television production and among documentary and experimental filmmakers, these and other advances in tape technology enabled ¼-inch to supersede 35mm magnetic film for many applications in the mainstream film industry. Notable among developments was the introduction of a technique invented by the physicist Ray Dolby which significantly reduced noise levels on magnetic recordings. In crude terms, it worked by passing the input signal through an electronic device which increased the modulation level in weaker areas of the signal's dynamic range (i.e. boosted the recorded volume of quieter sounds), in order to raise the signal to noise ratio. In playback the process was reversed, thereby restoring the balance of the original input signal. The first commercial application for Dolby noise reduction was in the consumer audio market, first in LP mastering and then as an add-on device to the 'compact cassette' format, launched by the Dutch consumer electronics giant Philips in the early 1960s. Cassettes had originally been envisaged as a low cost, easy to use format for consumer recording and playback, designed specifically for office dictation and educational use.<sup>72</sup> Philips and its licensed manufacturers soon discovered that customers were actually using it to play music recordings (either made from radio broadcasts, purchased in the form of pre-recorded cassettes or copied from LP records), and Dolby therefore produced a version of his noise reduction technology that was specifically designed to reduce the 'tape hiss' which characterised the cassette as a lower-quality medium than either LPs (which could not be recorded at home) or ¼-inch tape (which was significantly more expensive). As with virtually all the other forms of sound technology covered in this chapter, Dolby noise reduction was appropriated by the film industry after its benefits in other sectors had been established. Following a series of experiments, the first major feature film to use Dolby in post-production audio mixing was *A Clockwork Orange* (1971, dir. Stanley Kubrick).<sup>73</sup>

This pattern repeated itself yet again with digital recording. Given the extent of the issues and techniques around it which are common to both moving images and audio, the 'd-word' will be covered in one fell swoop in chapter eight. As far as its application to audio is concerned, the technology derived primarily from the computing and IT industries. The use of computers to represent, store and manipulate sound as digital data on a commercial basis have their origins (as with Lee de Forest and electronic amplification) in the telephone industry, and experiments in the early 1960s to establish if digitising transmissions could significantly increase the 'bandwidth', or capacity, of long-distance telephone cables. The record industry began to take an interest in the mid-1970s, when computer processing power and magnetic data storage increased to the point at which the sound quality of digital audio became comparable to that of analogue tape. The first generation of studio digital recorders, however, were not suitable for film industry use. Their mechanisms were those of helical scan Umatic video cassette recorders (see chapter six), adapted to read and write digital audio data rather than an analogue video signal. At that time, a helical scan tape mechanism was the only method available of reading and writing the volume of data necessary for digital audio and in real time. They were many times more expensive than ¼-inch tape technology, could only be used in fixed installations (i.e. they were not portable) and were not perceived by film studios to offer a significant improvement in the final mixed soundtrack as it was delivered to cinemas. Although the compact disc made digital audio playback in the home available from 1983, sales of the new format did not start to rise significantly until the end of the decade, and it was around this time that the use of digital sound first appeared on the film industry's agenda. One landmark development was the launch of the Digital Audio Tape (DAT) format by Sony in 1987. It recorded two channels of sound in an almost identical data format (and therefore quality) to that of the compact disc, but was much smaller and more rugged (despite also being based on a helical scan mechanism) than any previous digital recorder. Its use followed exactly the opposite trajectory to that of the analogue cassette: although it was originally intended as a consumer format, it is believed that opposition from the music industry, based on fears that the format would mainly be used by consumers to produce digital 'clones' of compact discs, kept hardware and media prices high.<sup>74</sup> In the event, DAT began to be used extensively in the early to mid-1990s by broadcasters and filmmakers as an origination medium (the portability of the equipment made it ideal for location use), as an effective replacement for ¼-inch, even though subsequent mixing and editing was still generally done using analogue magnetic tape. During the latter half of the 1990s, personal computer and audio technology began to integrate with a vengeance, as processor speeds and hard disc capacity increased to the point at which audio could be captured and manipulated as easily, as quickly and with as much versatility as other forms of data, such as text in a wordprocessor. A proliferation of stand-alone, hard-disc based audio recorders entered the market at around the turn of the century, and by the time of writing hard discs were being used almost exclusively for recording and post-production mixing in the film industry. Due to concerns over the long-term preservation of digital data (see chapters seven and eight), however, high-

quality magnetic analogue master elements are still frequently made, especially for higher end production materials, as archival storage masters.

## Conclusion

With the one exception of analogue optical sound, perhaps the most startling conclusion which can be drawn from this narrative is that none of the technologies which have been successfully used to add synchronised sound to film-based moving images were initially developed specifically for that purpose. Furthermore, and in each and every case, there was a significant time lag (ranging from a few years in the case of the 'key moment' constituent technologies to over four decades in the case of stereo) between that initial development and/or commercial exploitation elsewhere, and its mainstream appropriation by filmmakers. Acoustic recording could perhaps be characterised as an exception to this rule – the embryonic Victorian audio-visual industry in general and Edison in particular had been trying to synchronise it with moving images almost since the word go – but, compared with the sound technologies which later arrived and stayed, it could not be described as successful. Recordings were only ever produced to accompany a tiny fraction of the total films released during this period (1895 to 1926 approximately), and the emergence of the 'classical' model of film production in the aftermath of World War One killed off acoustic sound completely. Of the technologies which facilitated the 'key moment', microphones and electrical sound amplification were initially developed by the telephone and radio industries, and electrical disc cutting by the radio and retail music industries. The one significant exception to this rule – optical sound – was only developed to the point of commercial viability at all because of the obvious shortcomings of the alternative (discs).

Stereo sound had both its developmental and its commercial origins in the retail music industry: Alan Blumlein worked for EMI (Britain's largest record company) at the time of his experiments, and stereo LPs went on sale to consumers almost 30 years before stereo sound was widely available to the majority of cinemagoers. Magnetic recording was an experimental cottage industry without any obvious application until the Nazis got hold of it, and was then primarily a broadcasting medium before the film industry took it up soon after World War Two. And digital audio spent the best part of two decades (about the same time as electrical analogue amplification) being developed by the telephone and music industries before the film studios started to take an interest.

Why, therefore, has this area of moving image technology depended so heavily on the products of other industries and in most cases waited a considerable length of time before starting to use them? Some of the answers can be found in the unique relationship between culture, technology and economics which cinema has always embodied. As with the MPPC experiment of making sales of equipment conditional on sales of film stock, Edison believed that by synchronising acoustic recordings to films he could boost the revenue-earning potential of both. When that proved not to be the case (i.e. feature-length 'classical' silent films became the economically

dominant format, with which his sound technology was incompatible), he fell by the wayside. A widespread myth exists to this day that Sam Warner decided to launch Vitaphone simply because his company was losing market share to the other Hollywood majors, and therefore 'gambled on a new technology'<sup>75</sup> in order to try and win it back. Though the extent of Warner Bros.' financial dire straits has certainly been exaggerated in many accounts of the 'key moment', the fact remains that they would not have introduced sound when they did either (i) in the absence of a perceived business case, or (ii) if their system was not significantly more compatible than its predecessor with the cultural mode of film production which had by then become a mainstay of the Hollywood film industry.

Magnetic, stereo and digital sound were also introduced on the same basis: when it became clearly apparent that their use would enhance the production of the sort of films which the film industry knew would make money at the box office, and not before. If these technologies had been used in other industries for many years previously, then so much the better: there was less of a chance of mistakenly adopting a system which would ultimately prove unsuitable (as had been the case with Vitaphone) and less of the initial research and development costs to absorb.

Sound, therefore, has been always been treated very differently in connection with film than in connection with other moving image technologies. There was never any such thing as 'silent' television broadcasting on any significant scale (television grew directly out of radio in an industrial sense, in which audio was the mainstream, and the 'key moment' had already happened in film by the time regular broadcasting began); no mainstream videotape recorder was ever built which did not record and reproduce at least one soundtrack in synchronisation with the picture (apart from those designed for CCTV recording); and in the world of digital media, as with broadcasting, audio essentially came first. With film, moving images came first, and continued to be the core business. The technology was made to 'work' on a commercial basis over thirty years before audio technology became available which complimented and enhanced its cultural and economic role. For the first quarter of its life, therefore, the film industry did not integrate sound technology as a key component in its cultural norms and its business model, and has never wholly depended on it. Indeed, the first generation of sound technology was found wanting, and duly discarded.

Unlike the key developments in other areas of film-based moving image technology (e.g. film emulsions, formats, cameras and colour), sound remained an accessory to its key business and culture, right through to the introduction of digital audio in the 1990s. Whereas developments within image technology either came from within the industry itself or from its allied service sectors (for example, film stock manufacturers), sound technology was taken off the shelves of other cultural, communications and mass-media industries as and when it was perceived to deliver a benefit, and its integration then enshrined through the use of standardisation. It is unlikely to have escaped your notice that I began this chapter by suggesting that film historians tend to dwell too much on the 'key moment' of the 1926–32 conversion to sound, and then proceeded to spend over half its total word-length doing precisely that. The 'key moment' is simply the most dramatic and wide-ranging example of

this phenomenon, which can also be identified in every other significant development in the history of film sound. It is hoped that its coverage in this chapter has demonstrated that fact, and thereby contextualised the role of sound in relation to the other constituent technologies of film-based moving images, the last of which will be discussed in the following chapter.