

## 2 'Pictures That Talk and Sing'

### Sound History and Technology

DAVID COOPER

Synchronization sometimes went disastrously wrong in early experiments to join sound and film, particularly if these two components were reproduced separately and got out of step, say due to the film having been damaged, a section removed and the ends joined together. Austin Lescarbourea, managing editor of *Scientific American* from 1919 to 1924, devoted a chapter of his book *Behind the Motion-Picture Screen* (first published in 1919) to the topic of 'Pictures That Talk and Sing' and described a scene from the film *Julius Caesar* (1913) – an early Shakespeare adaptation presented in the Kinetophone sound system with sound provided by synchronized Edison sound cylinders – in which an actor 'suddenly sheathed his sword, and a few seconds later came the commanding voice from the phonograph, somewhere behind the screen saying: "Sheathe thy sword, Brutus!"' (Lescarbourea 1921, 292). The audience's response was, understandably, one of hilarity.

Lescarbourea, with considerable prescience, foresaw that the principles involved in new photographic methods of recording sound – then still at an experimental stage – would 'some day form the basis of a commercial system' (*ibid.*, 300). Early attempts at 'sound on picture' (such as Lee de Forest's Phonofilm) suffered from problems of noise, but by 1928 many of these had been resolved and a soundtrack that had reasonable fidelity became available through Movietone, which, like Phonofilm, had the advantage of the soundtrack actually lying on the print beside the pictures, thus removing the problem of 'drift' between two separate, albeit connected, mechanisms. The fascinating period during which filmmakers came to terms, in various ways, with the implications of the new technology is charted in detail in Chapter 1 of the present volume.

According to John Michael Weaver, James G. Stewart (a soundman hired by RKO in 1931 and chief re-recording mixer for the company from 1933 to 1945) recalled that the first sound engineers in film wielded considerable power:

during the early days of Hollywood's conversion to sound, the production mixer's power on the set sometimes rivalled the director's. Recordists were able to insist that cameras be isolated in soundproof booths, and they even had the authority to

cut a scene-in-progress when they didn't like what they were hearing.

(Weaver 1993, 12)

Stewart noted that the impositions made by the sound crew in pursuit of audio quality in the first year of talkies had a detrimental effect on the movies, bringing about 'a static quality that's terrible'. Indeed, one might argue that film sound technology, from its very inception, has been driven forward by the attempt to accommodate the competing demands of sonic fidelity and naturalness.

Peter Copeland has remarked how rapidly many of the most important characteristics of sound editing were developed and adopted. By 1931, the following impressive list of techniques was available:

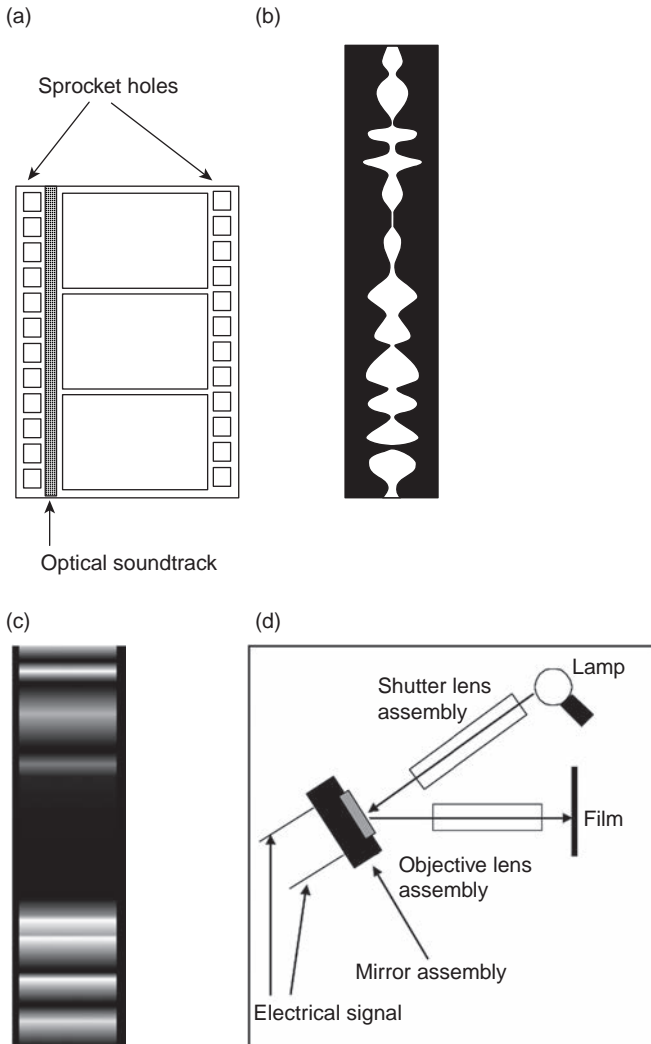
cut-and-splice sound editing; dubbing mute shots (i.e. providing library sounds for completely silent bits of film); quiet cameras; 'talkback' and other intercom systems; 'boom' microphones (so the mike could be placed over the actor's head and moved as necessary); equalization; track-bouncing; replacement of dialogue (including alternative languages); filtering (for removing traffic noise, wind noise, or simulating telephone conversations); busbars for routing controlled amounts of foldback or reverberation; three-track recording (music, effects, and dialogue, any of which could be changed as necessary); automatic volume limiters; and synchronous playback (for dance or mimed shots).

(Copeland 1991, 21)

Re-recording, which was used less frequently before 1931, would become increasingly prevalent through the 1930s due to the development of noise-reduction techniques, significantly impacting on scoring and music-editing activities (Jacobs 2012).

## The Principles of Optical Recording and Projection

A fundamental principle of sound recording is that of transduction, the conversion of energy from one form to another. Using a microphone, sound can be transformed from changes in air pressure to equivalent electrical variations that may then be recorded onto magnetic tape. Optical recording provides a further means of storing audio, either through the changes in density, or more commonly the changes in width of the developed section of a strip of photographic emulsion lying between the sprocket holes and the picture frames. Figure 2.1(a) displays one of a number of different formats for the variable-width track with a solid black edge on the left-hand side, which is termed the unilateral variable area, shown in Figure 2.1(b). Other alternatives (bilateral, duplex and push-pull variable areas) have solid edges on the right side or on both sides. In variable-density recording, the photographic emulsion can take on any state between undeveloped (black – no light passes



**Figure 2.1** 35 mm film with optical soundtrack (a), a close-up of a portion of the variable-area soundtrack (b), a close-up of a portion of the variable-density soundtrack (c) and a simplified schematic view of a variable-area recorder from the late 1930s (d), based on Figure 11 from L. E. Clark and John K. Hilliard, 'Types of Film Recording' (1938, 28). Note that there are four sprocket holes on each frame.

through) to fully developed (clear – all light passes through). Between these two states, different 'grey tones' transmit more or less light: see Figure 2.1(c). In digital optical recording tiny dots are recorded, like the pits on a CD.

The technology for variable-area recording, as refined in the 1930s, involves a complex assembly in which the motion of a mirror reflecting light onto the optical track by way of a series of lenses is controlled by the electrical signal generated by the microphone: see Figure 2.1(d). If no signal is present, the opaque and clear sections of film will have equal areas, and as the

sound level changes, so does the clear area. When the film is subsequently projected, light is shone through the optical track and is received by a photoelectric cell, a device which converts the illumination it receives back to an electrical signal which varies in proportion to the light impinging on it, and this signal is amplified and passed to the auditorium loudspeakers. The clear area of the soundtrack, between the two opaque sections shown in Figure 2.1(b), is prone to contamination by particles of dirt, grains of silver, abrasions and so on. These contaminants, which are randomly distributed on the film, result in noise which effectively reduces the available dynamic range: the quieter the recording (and thus the larger the clear area on the film soundtrack), the more foreign particles will be present and the greater will be the relative level of noise compared to recorded sound. Much effort was expended by the studios to counteract this deficiency in optical recording by means of noise reduction in the 1930s, the so-called push-pull system for variable-area recording being developed by RCA and that for variable-width recording by Western Electric's ERPI (see Frayne 1976 and Jacobs 2012, 14–18).

Cinematography involves a kind of sampling, each 'sample' being a single picture or frame of film – three such frames are shown in Figure 2.1(a) – with twenty-four images typically being taken or projected every second. Time-lapse and high-speed photography will also use different frame rates. Early film often had much lower frame rates, for instance 16 frames per second (fps). The illusion of continuity generated when the frames are replayed at an appropriate rate is usually ascribed to persistence of vision (the tendency for an image to remain on the retina for a short time after its stimulus has disappeared), though this is disputed by some psychologists as an explanation of the phenomenon of apparent or stroboscopic movement found in film: as Julian Hochberg notes, 'persistence would result in the superposition of the successive views' (1987, 604). A major technical issue which the first film engineers had to confront was how to create a mechanism which could produce the stop-go motion needed to shoot or project film, for the negative or print must be held still for the brief time needed to expose or display it, and then moved to the next frame in about 1/80th of a second. Attempts to solve this intermittent-motion problem resulted in such curious devices as the 'drunken screw' and the 'beater movement', but generally a ratchet and claw mechanism is found in the film camera.

Some film projectors still use one of the most ingenious of the early inventions, modelled on the 'Geneva' movement of some Swiss watches and called the 'Maltese Cross movement' because of the shape of its star wheel: see Figure 2.2. The wheel on the right has a pin (a) and a raised cam with a cut-out (b). As this cam-wheel rotates on its central pivot, the

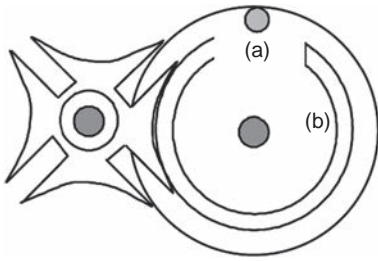


Figure 2.2 The Maltese Cross or Geneva mechanism.

star wheel on the left (which controls the motion of the film) remains stationary until its wings reach the cut-out and the gap between the wings locks onto the pin. The star wheel is now forced to turn by a quarter of a rotation, advancing the film by one frame. Unlike the intermittent motion required for filming and projection, sound recording and replay needs constant, regular movement. A pair of damping rollers (rather like a miniature old-fashioned washing mangle) therefore smoothes the film's progress before it reaches the exciter lamp and the photoelectric cell, which generate the electrical signal from the optical track and hence to the audio amplifier and the auditorium loudspeakers. As the soundtrack reaches the sound head some twenty frames after the picture reaches the projector, sound must be offset ahead of its associated picture by twenty frames to compensate.

## Soundtracks on Release Prints

Early sound films only provided a monophonic soundtrack for reproduction, and this necessarily limited the extent to which a 'three-dimensional' sound-space could be simulated. This was an issue of considerable importance to sound engineers. Kenneth Lambert, in his chapter 'Re-recording and Preparation for Release' in the 1938 professional instructional manual *Motion Picture Sound Engineering*, remarks:

Most of us hear with two ears. Present recording systems hear as if only with one, and adjustments of quality must be made electrically or acoustically to simulate the effect of two ears. Cover one ear and note how voices become more masked by surrounding noises. This effect is overcome in recording, partially by the use of a somewhat directional microphone which discriminates against sound coming from the back and the sides of the microphone, and partially by placing the microphone a little closer to the actor than we normally should have our ears.

(K. Lambert 1938, 76)

Whilst it was possible to suggest the position of a sound-source in terms of its apparent distance from the viewer by adjustment of the relative levels of

the various sonic components, microphone position and equalization, film-makers could not provide explicit lateral directional information. If, for instance, a car was seen to drive across the visual foreground from left to right, the level of the sound effect of the car's engine might be increased as it approached the centre of the frame and decreased as it moved to the right, but this hardly provided a fully convincing accompaniment to the motion. In order to create directional cues for sound effects, further soundtracks were required. Lest there be any confusion, the term 'stereo' has not usually implied two-channel stereophony in film sound reproduction, as it does in the world of audio hi-fi (the Greek root *stereos* (στερεός) actually means 'solid' or 'firm'); rather, it normally indicates a minimum of three channels. As early as the mid-1930s, Bell Laboratories were working on stereophonic reproduction (though, interestingly, no mention is made of these developments in *Motion Picture Sound Engineering* in 1938), and between 1939 and 1940 Warner Bros. released *Four Wives* and *Santa Fe Trail*, both of which used the three-channel Vitasound system, had scores by Max Steiner and were directed by Michael Curtiz (Bordwell *et al.* 1985, 359).

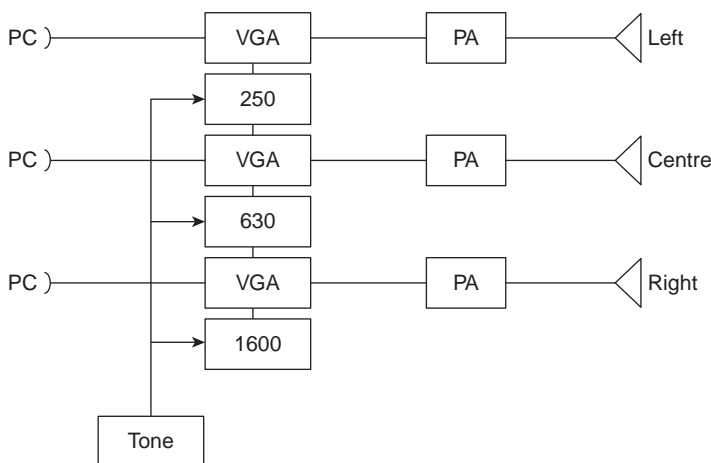
Disney's animation *Fantasia* (1940), one of the most technically innovative films of its time, made use of a high-quality multi-channel system called Fantasound. This system represented Disney's response to a number of perceived shortcomings of contemporaneous film sound, as was explained in a 1941 trade journal:

- (a) Limited Volume Range. – The limited volume range of conventional recordings is reasonably satisfactory for the reproduction of ordinary dialog and incidental music, under average theater conditions. However, symphonic music and dramatic effects are noticeably impaired by excessive ground-noise and amplitude distortion.
- (b) Point-Source of Sound. – A point-source of sound has certain advantages for monaural dialog reproduction with action confined to the center of the screen, but music and effects suffer from a form of acoustic phase distortion that is absent when the sound comes from a broad source.
- (c) Fixed Localization of the Sound-Source at Screen Center. – The limitations of single-channel dialog have forced the development of a camera and cutting technic built around action at the center of the screen, or more strictly, the center of the conventional high-frequency horn. A three-channel system, allowing localization away from screen center, removes this single-channel limitation, and this increases the flexibility of the sound medium.
- (d) Fixed Source of Sound. – In live entertainment practically all sound-sources are fixed in space. Any movements that do occur, occur slowly. It has been found that by artificially causing the source of sound to move rapidly in space the result can be highly dramatic and desirable.

(Garity and Hawkins 1941, 128)

In its final incarnation, the Fantasound system had two separate 35 mm prints, one of which carried four optical soundtracks (left, right, centre and a control track). Nine optical tracks were utilized for the original multitrack orchestral recordings (six for different sections of the orchestra, one for ambience, one containing a monophonic mix and one which was to be used as a guide track by the animators). These were mixed down to three surround soundtracks, the playback system requiring two projectors, one for the picture (which had a normal mono optical soundtrack as a backup) and a second to play the three optical stereo tracks. The fourth track provided what would now be regarded as a type of automated fader control. Tones of three different frequencies (250 Hz, 630 Hz and 1600 Hz) were recorded, their levels varying in proportion to the required levels of each of the three soundtracks (which were originally recorded at close to maximum level), and these signals were subsequently band-pass filtered to recover the original tones and passed to variable-gain amplifiers which controlled the output level: see Figure 2.3 for a simplified diagram. The cinema set-up involved three horns at the front (left, centre and right) and two at the rear, the latter being automated to either supplement or replace the signals being fed to the front left and right speakers.

As part of the research and development of the sound system, Disney engineers also invented the 'Panpot', or panoramic potentiometer, a device to balance the relative level of sounds in the left, centre and right channels that has become a fundamental control found in every modern mixing console. Using the Panpot, engineers could 'steer' sound across the sound-space, allowing them to produce the kind of lateral directional information discussed above. Given a cost of around \$85,000 to equip the auditorium, it



**Figure 2.3** A simplified block diagram of the Fantasound system (after Figure 2 in Garity and Hawkins 1941). PC indicates photocell, PA power amplifier and VGA variable-gain amplifier.

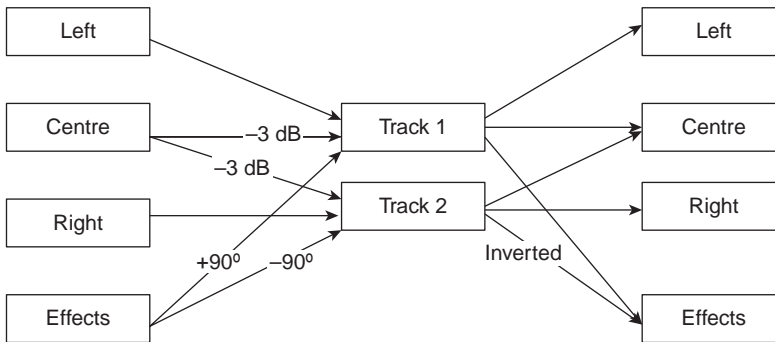
is perhaps unsurprising that only two American movie houses were willing to invest in the technology required to show *Fantasia* in the Fantasound version, although eight reduced systems, each costing \$45,000, were constructed for the touring road shows which used a modified mix of orchestra, chorus and soloists respectively on the three stereo tracks (see Blake 1984a and 1984b; Klapholz 1991).

Stereo sound in film was intimately tied to image size, and in CinemaScope – Twentieth Century-Fox's 2.35:1 widescreen format which dominated 35 mm film projection from the early 1950s – the soundtrack complemented the greater dramatic potential of the picture. The numbers 2.35:1 indicate the aspect ratio of the width and height of the projected picture, that is, the image width is 2.35 times the image height. (The Academy ratio which prevailed before widescreen films became popular is 1.33:1.) CinemaScope was not the first widescreen format: Cinerama, which was used for the best part of a decade from 1952, had a seven-track analogue magnetic soundtrack with left, mid-left, centre, mid-right and right front channels, and left and right surround channels.

CinemaScope films have four audio tracks, one each for right, centre and left speakers at the front of the auditorium, and one for effects. These are recorded magnetically rather than optically, with magnetic stripes between the sprocket holes and the picture, and between the sprocket holes and film edge. (A 70 mm format called Todd-AO, which had six magnetic soundtracks, was also available.) *The Robe*, Henry Koster's Biblical spectacle of 1953 with a score by Alfred Newman, was the first film to be released by Fox in the CinemaScope format. Although around 80 per cent of American cinemas were able to show films in CinemaScope by 1956, only 20 per cent had purchased the audio equipment required to reproduce the magnetic tracks, and viewers had the mixed blessing of a superior widescreen image and an inferior monophonic optical soundtrack. Theatre managers were reluctant to adopt magnetic reproduction, not merely because of the up-front cost of installing the equipment, but also because of the much greater hire charge for magnetic prints, which were at least twice as expensive as their optical equivalents.

A revolution in film-sound reproduction occurred in the 1970s with the development by Dolby Laboratories of a four-channel optical format with vastly improved sound quality. Dolby had found an ingenious method of fitting the three main channels and the effects channel into two optical tracks, using 'phase steering' techniques developed for quadraphonic hi-fi. Ken Russell's critically pilloried *Lisztomania* of 1975 was the first film to be released using the new format called Dolby Stereo, and it was followed in 1977 by George Lucas's *Star Wars* (later renamed *Star Wars Episode IV: A New Hope*) and Steven Spielberg's *Close Encounters of the Third Kind*,





**Figure 2.4** A simplified diagram of the Dolby Stereo system. On the left side are the four source channels, in the centre are the two film soundtracks derived from them and on the right are the four output channels played in the cinema.

both of which made great use of the potential offered by the system. Dolby Stereo involves adding the centre channel (dialogue, generally) equally to both left and right tracks, but at a lower level (-3 dB), the surround channel also being sent to both tracks, but 180° degrees out of phase relative to each other (on one the surround signal is +90°, on the other it is -90°). A matrixing unit reconstructs the four channels for replay through the cinema's sound-reproduction system: see Figure 2.4.

Dolby Laboratories improved the frequency response and dynamic range of the analogue optical audio tracks in 1987 by introducing their 'SR' noise-reduction process, and in 1992 a digital optical system with six channels was made available, which had the standard three front channels, one for low-frequency effects and two separate left and right rear channels for surround sound. Dolby's consumer formats, which have similar specifications to their cinema-sound products, have seen wide-scale adoption, though competing multichannel digital formats include DTS (Digital Theater Sound) which, in an approach that looks back to Vitaphone, uses six-channel sound stored on external media and synchronized to the film by timecode, and Sony's SDDS (Sony Dynamic Digital Sound), which employs the ATRAC (Adaptive Transform Acoustic Coding) compression technology designed for MiniDisc.

The technical improvements in sound quality have required cinema owners to invest in their projection and reproduction equipment if they wish to replay multichannel film sound adequately. A standard for such equipment, and the acoustics of the theatres which use it, was established by Lucas's company Lucasfilm, certified cinemas being permitted to display the THX logo, and required to apply annually for re-certification. The standard covers such diverse issues as background noise, isolation and reverberation, customer viewing angle, equipment and equipment

installation. THX can be seen, perhaps, as an attempt to complete the democratization of the cinema, for unlike the picture palaces of days gone by, every seat in the house should provide a similar (though not identical) hi-fidelity aural experience, which ideally relates to that of a soundtrack reproduced on a digital format with high-quality headphones.

In late 1998, agreement was reached between Dolby Laboratories and Lucasfilm THX for the introduction of a 6.1 surround-sound format called Dolby Digital-Surround EX, which includes a centre-rear speaker for more accurate placement of sound at the rear of the auditorium. The format, initially employed in 1999 on the first film in Lucas's second *Star Wars* trilogy, Episode I: *The Phantom Menace*, brought together two of the major players in film sound in an attempt to improve the sonic quality of the cinematic experience even further.

Cinema sound has continued to be a major focus for technical development, including the introduction of further digital formats such as Dolby Surround 7.1 (which employs a pair of surround channels on each side) in the release of *Toy Story 3* (2010); Barco Auro 11.1 (which places the 5.1 system on two vertical levels with an additional overhead channel) in *Red Tails* (2012); and Dolby Atmos full-range surround with up to sixty-four channels on separate levels in *Brave* (2012).

## Soundtracks at the Post-Production Stage

The soundtrack of the final release of a film is mixed from three basic types of audio material: dialogue, music and effects. These three elements (known as 'stems') are usually archived on separate tracks by the production company for commercial as much as artistic reasons: many films are also released in foreign-language versions and it would be expensive to re-record music and effects as well as dialogue, which is added by actors in an ADR (automatic dialogue replacement) suite as they watch a looped section of the film. The quantity of sound actually recorded on set or on location and used on the final soundtrack varies between productions, partly for pragmatic and partly for aesthetic reasons. At one extreme, all dialogue and effects may be added during post-production (after the main filming has been completed), and at the other, only location sound may be used. Although one of the first films to have a soundtrack created entirely in post-production was Disney's Mickey Mouse short *Steamboat Willie* (1928), the transition to this being the norm for most live-action sound films took another decade (K. Lambert 1938, 69). If the film is recorded on a sound stage in a studio, much of the dialogue will probably be of acceptable quality, whereas audio collected on location may suffer from

background noise, rendering it unusable. For example, in Ron Howard's *Willow* (1988), 85 per cent of the dialogue was added in post-production, whereas in Spielberg's *Indiana Jones and the Last Crusade* (1989), only 25 per cent was (Pasquariello 1993, 59). James Monaco notes that in some instances Federico Fellini had not even written the dialogue for his films until shooting had finished (1981, 106).

Sound-effects tracks can be subdivided into three main categories: spot effects, atmospheres and Foley. Spot (or hard) effects may be specially recorded or taken from effects libraries on CD or other digital media, and include most discrete sounds of relatively short duration: dogs barking, engines starting, telephones ringing and so on. Atmospheres (or wild tracks, so called because they are non-synchronous with the action) are of longer duration than spots (though a hard-and-fast line between the two types can be hard to draw), may be recorded on set or specially designed and are intended to enhance the aural impression of a particular space or location. Foley effects are named after Jack Foley, who pioneered the recreation of realistic ambient sounds in post-production at Universal in the late 1920s. Foley artists work on a specially designed sound-proofed and acoustically dead recording stage where they watch the film on a monitor and perform the sound effects in sync with the action to complement or replace the background sounds recorded on set. Creature sounds, many of which must be invented to supply the language of extinct or fantastic animals (for example, the dinosaurs of Spielberg's 1993 film *Jurassic Park*, or the aliens in *Star Wars*), form a further sub-category of effects. These may derive from real animal sounds which are recorded and processed or superimposed – for example, Chewbacca in *Star Wars* is in part a recording of walrus.

An alternative strategy for dealing with effects tracks is to consider them in terms of the apparent location of their constituent elements in a flexible three-dimensional sound-space. Thus it is possible to discriminate between a sonic foreground, middleground and background that can dynamically mirror the equivalent visual planes (irrespective of whether the sources are in view). The accurate construction of these sound-spaces became more significant with the development of sound systems like Dolby Stereo and Dolby Digital in which finer detail could be distinguished, and the re-equipping of cinemas with high-quality reproduction equipment according to the THX standard. Sound designers such as Alan R. Splet, who worked closely with director David Lynch on his films from *Eraserhead* (1976) to *Blue Velvet* (1986), showed just how creative an area this can be, for their work often went beyond the naturalistic reproduction of sound-spaces, at times adopting a role analogous to that of nondiegetic music. Lynch's own telling remark that 'people call me a director, but

I really think of myself as a sound-man' underlines this trend to place the sonic elements on an equal footing to that of the visuals (Chion 1995b, 169).

During the dub, a final stereo mix is produced from the dialogue, effects (including atmospheres and Foley) and music tracks, and this is used to generate the soundtracks on the release print. The dub is normally a two-stage process, the first of which involves a pre-mix in which the number of tracks is reduced by combining the many individual effects and dialogue tracks into composite ones. In *Indiana Jones and the Last Crusade*, for instance, twenty-two tracks were sometimes used for Foley alone; for the television show *Northern Exposure* (CBS, 1990–95), up to sixty-nine analogue tracks held music, dialogue and effects; and twenty-eight eight-track recorders were used for effects on Roland Emmerich's 1998 blockbuster *Godzilla* (James 1998, 86). The pre-post-prodmix results in individual mono or stereo tracks (four or six, depending on the system) for dialogue and effects (Foley, spots and atmospheres), and these are mixed with the music tracks by three re-recording mixers at the final dub to generate the release soundtrack. A track or group of tracks called 'music and effects' (M&E) is also mixed from the individual music and effects tracks to retain decisions made by the director and editors about the relative levels of the individual components, for the reason explained above.

## Synchronization

In every case, the soundtrack must be locked or synchronized to picture, and this has provided problems since the earliest sound films with a score. For dialogue and effects tracks it is largely a matter of matching a visual image with its sonic correlate (for example, the sound of gunfire as the trigger of a gun is pressed, or engine noise as a car's ignition switch is turned). Often there will be an unambiguous one-to-one mapping between image and sound, and difficult as the operation may be for the editor, there is a 'correct' solution. This is not necessarily the case with music, for although the technique of tight synchronization between image and music (mickey-mousing, otherwise known as 'catching the action') is frequently found, there will generally be a much subtler relationship between the two.

Steiner has often been credited with the invention of the click track, as an aid to his work on the score to John Ford's *The Informer* (1935), though in his unreliable and unpublished memoirs he makes various unsubstantiated claims about his invention or use of the click track for earlier RKO films produced in 1931 and 1932 (Steiner 1963–4). It seems that Disney's

crew had used a similar device some time earlier, to assist the synchronization of animated film and music, filing patents in 1928, 1930 and 1931. In *Underscore*, a 'combination method-text-treatise on scoring music for motion picture films or T.V.' written by the prolific composer Frank Skinner in 1950, the main uses of the click track in the classic Hollywood film were enumerated:

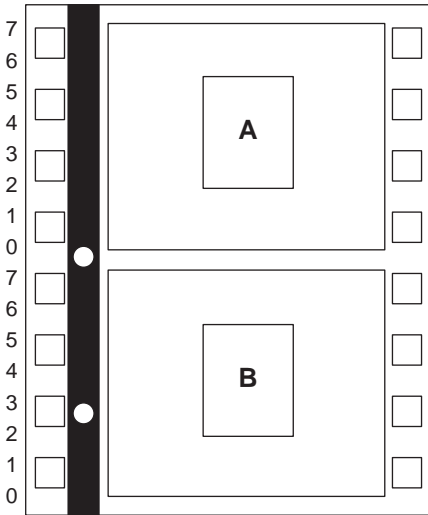
- 1 To catch intricate cues on the screen by writing accents to fall on the corresponding clicks, which will hit the cues automatically.
- 2 To speed up, or make recording easier, if the music is to be in a steady tempo. This eliminates any variation of tempo by the conductor.
- 3 To record the correct tempo to dancing, such as a café scene which has been shot to a temporary tempo track to be replaced later. The cutter makes the click track from the original tempo track.
- 4 To lead into a rhythmic piece of music already recorded. The click track is made in the correct tempo and the number of bars can be adjusted to fit.
- 5 To insert an interlude between two pieces of music, if rhythmic, to fit perfectly in the correct tempo.
- 6 To record any scene that has many cues and is difficult for the conductor. The music does not have to be all rhythmic to require the use of a click track. The composer can write in a rubato manner to the steady tempo of clicks.

(Skinner 1950, iv)

Although the technology used to generate the click track may have changed over the years, its function is still much the same for scores using live musicians.

The click track produces an audible metronomic tick calibrated to the prevailing tempo at any point in the score, and this is fed to the conductor (and normally the other musicians) by means of a distribution amplifier and headphones. In its earliest form, it required holes to be punched in the undeveloped (opaque) optical track on a film print. As film normally travels at 24 fps, if a hole is punched in the optical track next to the first sprocket hole of every twelfth frame, an audible click is produced each time a hole passes the sound head (every half a second), producing a metronome rate of 120 beats per minute. In order to provide a greater range of possible tempi, holes can be punched in any of eight different positions on a frame – next to a sprocket hole, or between sprocket holes (see Figure 2.5). Thus, for example, to set a tempo to 128 beats per minute, a hole is punched every 11 frames.

Given that each 1/8th of a frame of film represents approximately 0.0052 seconds, it is possible to calculate the length of time, and thus film, that a crotchet will take at any metronome mark. For example, if the metronome mark is 85, each crotchet will last for 60/85 seconds (around 0.706 seconds). Dividing this latter figure by 0.0052 gives just



**Figure 2.5** Punched holes in the optical soundtrack. The punch in frame A is at the top of the frame, the punch in frame B is at the 3/8 frame position. Numbers at the left-hand side of the frame indicate the frame position for punches. (Note that the film is travelling downwards, so the numbering is from the bottom of the frame upwards.)

over 135, and dividing again by 8 (to find the number of frames) produces the figure 16<sup>th</sup> frames. *Accelerandi* and *rallentandi* may also be produced by incrementally decreasing or increasing the number of 1/8th frames between punches, although this can require some fairly complicated mathematics. For many years, film composers have used 'click books' such as the Knudson book (named after the music editor Carroll Knudson, who compiled the first such volume in 1965), which relate click rates and click positions to timings, allowing the estimation of where any beat will lie in minutes, seconds and hundredths of a second.

As well as the audio cues provided by a click track, several forms of visual signal, known as streamers, punches and flutters, can be provided. Until the 1980s, streamers were physically scraped onto the film emulsion as diagonal lines running from left to right, over forty-eight to ninety-six frames (two to four seconds of film). Figure 2.6 illustrates a streamer running over nine frames – normally these would extend over three, four or five feet of film, and the line would look less oblique than it does here. When projected, during which process the film runs downwards and the image is inverted, a streamer appears as a vertical line progressing from the left to the right of the screen, terminated by a punched hole in the film (roughly the diameter of a pencil) which displays as a bright flash. Flutters offer a subtler and looser means of synchronization than the click track. Here clusters of punches in groups of three, five or seven are placed on a downbeat or at the

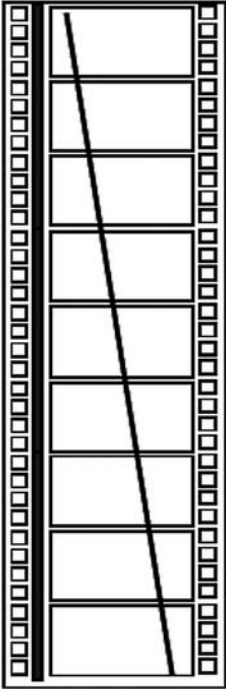


Figure 2.6 Schematic illustration of the visual cue known as a streamer, shown running over nine frames.

beginning of a musical cue, each punched frame having an unpunched neighbour (the pattern of punched [o] and unpunched [-] frames can be imagined as forming one of the following sequences: o-o-o, o-o-o-o-o or o-o-o-o-o-o-o). This method of 'fluttering' punches was devised by prolific film composer and studio music director Alfred Newman, and has been called the Newman system in his honour.

Almost invariably in contemporary film work, some form of computer-based system will generate the click track and visual cues in conjunction with a digital stopwatch on the conductor's desk. Auricle: The Film Composer's Time Processor, released in 1985 by the brothers Richard and Ron Grant, combined most of the time-based facilities required by the composer, including streamers in up to eight colours, allowing differentiation between varying types of cue. More recently, software applications such as Figure 53's Streamers have been widely adopted by composers and studios.

For many film composers, the real significance of such software is its ability to calculate appropriate tempi to match hit points. These are visual elements to be emphasized, often by falling on a strong beat or through some other form of musical accentuation. It is not difficult to calculate a fitting tempo and metre when there are only two such hit points, but the

maths can become considerably more complicated when there are three or more irregular ones. Appropriate software permits the composer to construct the rhythmic and metrical skeleton of a cue rapidly.

## Time Code

Much of the functionality of such software is premised on SMPTE/EBU (Society of Motion Picture and Television Engineers/European Broadcasting Union) time code. This offers a convenient way of labelling every frame of video by reference to its position in terms of hours, minutes, seconds and frames (hh:mm:ss:ff) relative to a start-point, and is available in two basic forms: vertical interval time code (VITC) and linear or longitudinal time code (LTC). There are a number of different and incompatible standards for time-code frame-rate: 24 fps for film, 25 fps for British and European video (PAL/SECAM) and 29.97 and 30 fps for American and Japanese video (NTSC).

Frame rates for video are related to the mains frequency. In the United Kingdom this is 50 Hz, and the frame rate is set to half this value (25 fps) whereas in the United States, which has a 60 Hz mains frequency, 30 fps was originally established as the default rate. With the development of colour television, more bandwidth was required for the colour information, and so the frame rate was reduced very slightly to 29.97 fps. Given that there are only 0.03 frames fewer per second for 29.97 fps than for 30 fps, this appears a tiny discrepancy, but over an hour it adds up to a very noticeable 108 frames or 3½ seconds. Thus 'drop frame' code was developed: the counter displays 30 fps, but on every minute except 0, 10, 20, 30, 40 and 50, two frames are dropped from the count (which starts from frame 2 rather than 0).

To accommodate these different rates, time code must be capable of being written and read in any version by a generator. In each case, a twenty-four hour clock applies, the time returning to 00:00:00:00 after the final frame of twenty-three hours, fifty-nine minutes and fifty-nine seconds. By convention, time code is not recorded to start from 00:00:00:00, but shortly before the one-hour position (often 00:59:30:00 to allow a little 'pre-roll' before the start of the material) to avoid the media being rolled back before the zero position, and confusing the time-code reader.

SMPTE/EBU time code is found in every stage of production and post-production, except in initial photography on analogue film, for it is not possible to record it on this medium; instead, 'edge numbers' sometimes called Keycode (a trademark of Kodak) are used, marked at regular intervals during the manufacturing process in the form of a barcode. Most film editing



is now digital and takes place on computer-based systems, which have almost entirely replaced the old flatbed 16 mm and 35 mm editing tables such as the Steenbeck (which in turn ousted the horizontal Moviola). Computer-based editing is described as a non-linear process, because there is no necessary relationship between the relative positions of the digital encoding of the discrete images and their storage location, which is dependent upon the operating system's method of file management. An important aspect of editing performed on a non-linear system, whether the medium is video or audio, is that it is potentially non-destructive: instead of physically cutting and splicing the film print or magnetic tape, the operator generates edits without affecting the source data held in the filestore. This will normally involve the generation of some kind of play list or index of start and stop times for individual takes which can be used by the computer program to determine how data is to be accessed.

Sound, picture and code are transferred to the computer where editing takes place and an edit decision list (EDL) is created, which indicates the start and end time-code positions of every source section of the picture or sound and their position in the composite master. Where analogue film is still used, this list (see Table 2.1 for an example) is employed to cut the original negative from which the projection print will finally be made once its time-code values have been converted back to Keycode frame positions.

Composers now generally work away from film studios, often preparing the complete music track in their own studios using a version of the film in a digitized format to which a MIDI-based sequencer is synchronized. Functionality provided by sequencers has become increasingly sophisticated and most not only record and play MIDI data, but also digital-audio data files recorded from a live source such as an acoustic instrument or generated electronically. The integration of MIDI and digital recording and editing has greatly simplified the problems of synchronization for the composer and

Table 2.1 *A section of an edit decision list (EDL).*

V and A indicate video and audio, respectively. V1, V2, A1 and A2 are four different source tapes of video and audio.

Cue	Type	Source	Source in	Source out	Master in	Master out
1	V	V1	06:03:05:00	06:04:09:21	00:01:00:00	00:02:04:21
2	A	A1	01:10:01:05	01:11:03:20	00:01:00:00	00:02:02:15
3	V	V2	13:20:05:00	13:20:55:00	00:02:04:21	00:02:54:21
4	A	A2	01:22:00:12	01:22:05:22	00:02:29:00	00:02:34:10
5	A	A1	06:05:10:03	06:05:29:05	00:02:35:01	00:02:54:03

sound editor, for interactions between the two sources of musical material can be seamlessly controlled. This is particularly relevant to the fine-tuning of cues, especially when the composed music is slightly longer or shorter than required. In the analogue period there were two main methods of altering the length of a cue: by physically removing or inserting sections of tape (possibly causing audible glitches) or by slightly speeding it up or slowing it down. In the latter case, this produces a change of pitch as well as altering the duration: faster tape speed will cause a rise of pitch, slower speed a fall. This may be reasonable for very small changes, perhaps a sixty-second cue being shortened or extended by one second, but larger modifications will become audibly unacceptable; for instance, shortening a one-minute cue by three seconds will cause the pitch to rise by almost a semitone, and produce a marked change in timbre.

MIDI sequences preclude this problem, for the program can simply recalculate note lengths according to the prevailing tempo, and this will cause the cue to be appropriately shortened or lengthened. Digital signal-processing techniques offer an equivalent method of altering digital-audio cues called time stretching, which does not result in a change of pitch. At one time such procedures were only available on expensive equipment in research institutes and large studios, but the enormous increase in the power of computer processors and reduction in the cost of computer memory and hard-disk space has made them readily accessible to the composer working on a personal computer or hardware sampler.

## Music and the Processes of Production

It is customary to consider the making of a film as being divided into three phases called pre-production, production and post-production. Depending on the nature of the film, the composition and recording of musical components may take place in any or all the phases, though the majority of the musical activities will usually take place in post-production. If music is scored or recorded during pre-production (generally called pre-scoring or pre-recording), it will almost certainly be because the film requires some of the cast to mime to an audio replay during shooting, or for some other compelling diegetic reason. Oliver Stone's *The Doors* (1991), a movie about Jim Morrison, is a good example of a film with a pre-recorded soundtrack, some of it being original Doors material and the rest being performed by Val Kilmer, who plays the lead role (Kenny 1993). Production recording of music has become generally less common than it was in the early sound films, largely because of the problems associated with the recording of hi-fi sound on location or on sets not optimized for

the purpose. Bertrand Tavernier's jazz movie *'Round Midnight* (1986), with a score by Herbie Hancock, is a somewhat unusual example, being shot on a set specially built on the scoring stage where orchestral recordings normally take place (Karlin and Wright 1990, 343).

For the majority of films, most of the nondiegetic music is composed, orchestrated, recorded and edited during post-production. By the time the composer becomes involved in the process, the film is usually already in, or approaching its final visual form, and thus a relatively brief period is set aside for these activities. Occasionally composers may be involved much earlier in the process, for example David Arnold was given the film for the 1997 Bond movie *Tomorrow Never Dies* reel by reel during production (May to mid-November), very little time having been allocated to post-production. Karlin and Wright indicate that, in 1990, the average time allowed from spotting (where the composer is shown the rough cut and decides, with the director, where the musical cues will happen) to recording was between three and six weeks. James Horner's 1986 score for James Cameron's *Aliens*, which has 80 minutes of music in a film just short of 140 minutes in duration, was completed in only three and a half weeks, and such short time scales remain common. Ian Sapiro has modelled the overall processes associated with the construction of the score at the level of the cue in his study of Ilan Eshkeri's music for *Stardust* (2007) and this is illustrated in Figure 2.7 (Sapiro 2013, 70).

Directors often have strong feelings about the kind of music they want in their films, and their predilections will govern the choice of composer and musical style. As an aid during the film-editing process, and as a guide for the composer during spotting, temp tracks (i.e. temporary music tracks), which may be taken from another film or a commercial recording, and will have some of the characteristics they require, are often provided. In some cases, these temp tracks will form the final soundtrack, most famously in Stanley Kubrick's *2001: A Space Odyssey* (1968), for which Alex North had composed a score, only to have it rejected in favour of music by Richard Strauss, Johann Strauss and Ligeti (see Hubai 2012 for a detailed study of rejected scores). If a film is rough-cut to a temp track, it may take on aspects of the rhythmic structure of the music, limiting the composer's freedom of action.

Spotting results in a list of timings and a synopsis of the accompanying action called timing notes or breakdown notes (see Table 2.2) that are given to the composer. Like the edit decision list discussed above, these notes provide a level of detail which not only suggests the kind of musical codes to be employed (given that the majority of films and television programmes rely upon the existence of commonly accepted musical forms of signification), but also gives information about its temporal character (metre and rhythm). Cues are listed according to the reel of film in which they appear (each reel of film being 1000 feet or just over

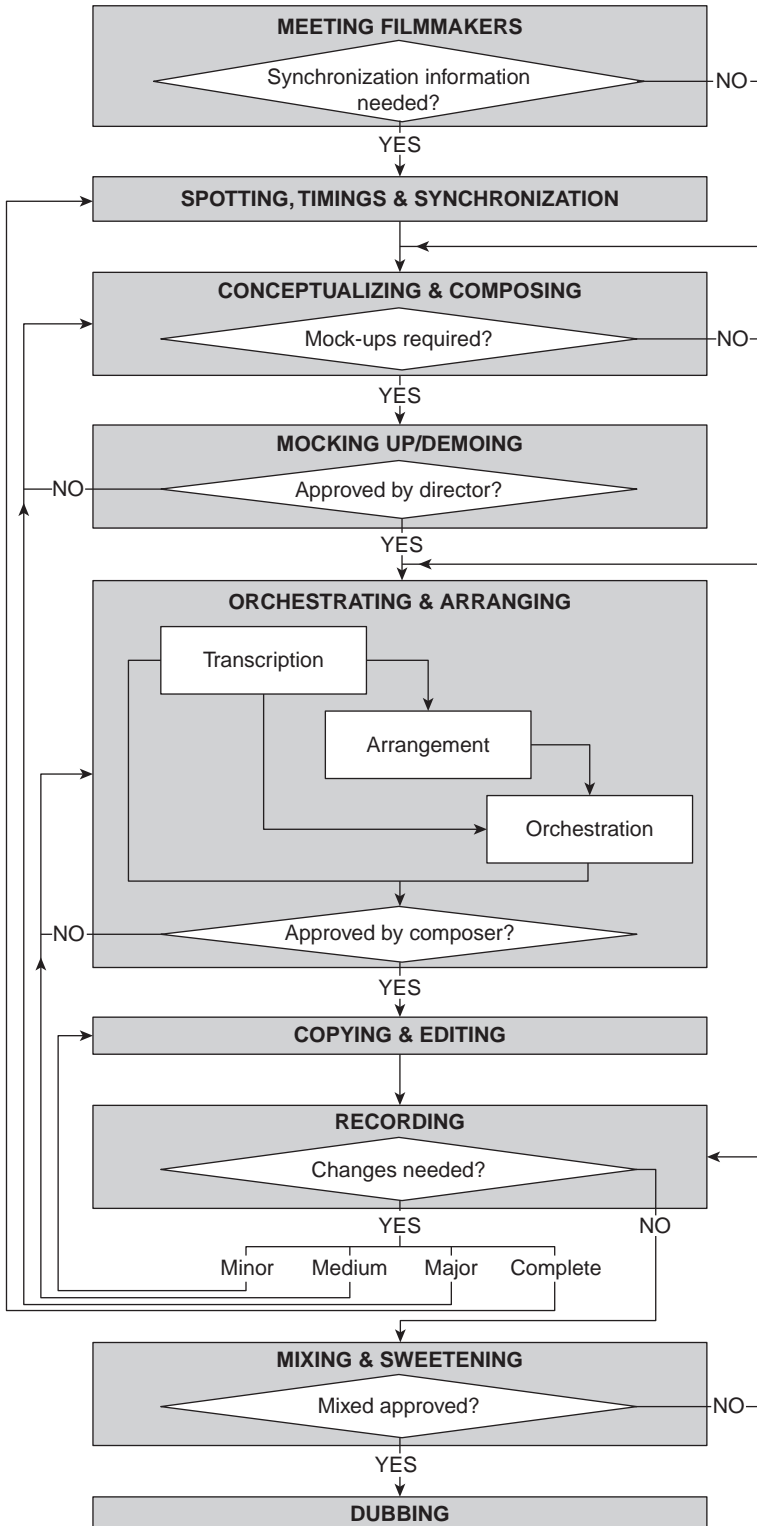


Figure 2.7 Ian Sapiro's model of the overall processes associated with film scoring as a decision matrix.

Table 2.2 *Timing sheet for cues in an imaginary film.*

2M2 refers to the cue number (second musical cue on the second reel of film). Time is total elapsed time from the beginning of the cue in minutes, seconds and hundredths of seconds.

Time code	Time	Action
		Night-time outside a large urban building. Eerie moonlight shines on wet road.
01:45:30:00	0:0.0	Music starts as boy and girl stare morosely into each other's eyes.
01:45:44:05	0:14.21	Ambulance passes with siren sounding.
01:46:01:00	0:31.00	Man emerges from office doorway.
01:46:02:20	0:32.83	Pulls out his large leather wallet.
01:46:08:23	0:38.96	Gives boy £20.00 note and takes package.
01:46:44:00	1:14.00	End of cue.

eleven minutes), are identified as being a musical cue by the letter M and are numbered according to their order in the reel.

As already implied, the 'factory' conditions of music production prevailing in the classic Hollywood system have partly given way to a home-working freelance ethos, with composers operating from their own studios. Certainly, high-budget films often still employ a large retinue of music staff including a music supervisor taking overall control, a composer, orchestrators, copyists and a music recordist, with orchestral musicians hired in for the recording sessions; but many lower-budget movies and television programmes have scores entirely produced by the composer using synthesized or sampled sounds ('gigastrated'), and returned to the production company on digital files ready for final mixdown. While this reduces the cost of the score dramatically, and lessens the assembly-line character of the operation, some of the artistic synergy of the great Hollywood partnerships of composers and orchestrators may have been lost. Although it may be theoretically possible, using a sequencer connected to sound libraries, to produce scores of the density and finesse of those written by Hollywood partnerships such as Max Steiner and Hugo Friedhofer, or John Williams and Herb Spencer, the short time frame for a single musician to produce a complex score can restrict creativity.

## New Technology, Old Processes

Since the 1980s, digital technology has infiltrated almost every aspect of the production and reproduction of film, and in the second decade of the

twenty-first century the major US studios such as 20th Century Fox, Disney and Paramount moved entirely to digital distribution. Despite this, the processes that underpin filmmaking have remained remarkably stable, and although pictures and TV shows may no longer be released on physical reels of celluloid, many of the concepts and much of the terminology from the analogue period retain currency. Equally, while the function of film music has been more elaborately theorized concurrently with the period of transition from analogue to digital, the role of the composer in US and European mainstream film and cinema, and of his or her music, has changed less than might have been expected in this technologically fecund new environment. While there certainly have been some radical developments, particularly at the interface between music and effects, the influence of 'classical Hollywood' can still be strongly detected in US and European mainstream cinema, an indication, perhaps, of both the strength of those communicative models and the innate conservatism of the industry.