Misusing Sights as Sounds:
the Infringements of Radio Drama
on the
Making of CITIZEN KANE

Evan Wm. Cameron

Professor Emeritus
Senior Scholar in Screenwriting

Graduate Programmes,
Film & Video and Philosophy

York University

[Presented in class from 1968. Recorded in the fall of 1971 as one of "Nine Lectures on Film Design" under the title "CITIZEN KANE and the Effects of Radio on the Coming of Sound to the American Film" by WBUR FM, Boston, Massachusetts, for broadcast in the spring of 1972; rebroadcast in the spring of 1973 on WROR and WRKO, Boston, and distributed thereafter through the National Public Radio network and local stations. A revision under the same title was presented on 05 April 1977 to the Conference on Radio Literature, University of Durham (England), the text appearing a year later as pages 85-98 of the *Proceedings of the Conference on Radio Literature* (Durham: University of Durham, 1978). A further revision entitled "CITIZEN KANE: the Influence of Radio Drama on Cinematic Design" appeared in 1980 as pages 202-216 of *Sound & the Cinema: The Coming of Sound to American Film* (New York, New York: Redgrave Publishing Company, 1980), the anthology that I compiled, edited (with the assistance of William F. Wilbert and Joan Evans-Cameron) and introduced of contributions to a symposium at George Eastman House, Rochester, New York in October 1973 on "The Coming of Sound to the American Film". For additional information, see my Curriculum Vitae.]

Misusing Sights as Sounds:
the Infringements of Radio Drama
on the Making of CITIZEN KANT

> ... I spent a year cutting it [CITIZEN KANE] and saw it 7000 times, but never since. I can't bear to see myself up there; I keep noticing things I want to change. When I made KANE, I had a sort of innocent assurance. I hadn't learnt what you aren't supposed to be able to do ... In handling a camera I feel I have no peer. But what De Sica can do, that I can't do. I ran his SHOESHINE again recently and the camera disappeared, the screen disappeared; it was just life ...
>
> Orson Welles[1]

Many arts have influenced the cinema over extended periods of time. One art – radio drama – is an exception, for we can date the onset of its influence from the coming of synchronous sound to the cinema in 1926 and the culmination of it with the creation of CITIZEN KANE in 1941. Film and radio drama were thereafter to part company, each having learned what it could from the other.

What had the cinema learned from radio drama? If we look closely at CITIZEN KANE, we can learn much about the virtues and limitations of radio design, cinematical design and the design of CITIZEN KANE itself. For CITIZEN KANE is a puzzling film to many and ought to be puzzling to many more. By common consent it is a great film, yet its greatness has been misunderstood. For three questions have seldom been faced, much less answered:

> Why, despite its reputation, has the design of CITIZEN KANE remained largely unemulated?
>
> > (Truffaut may have been right, though one wishes he had named names, when remarking that the film is "probably the one that has started the largest number of filmmakers on their careers"; but filmmakers, Welles and Truffaut included, have seldom *modelled* films after CITIZEN KANE.)[2]

The answer to the first question is implicit in the answers to a second and third:

---

[1] Quoted by Derick Grigs in his article "Orson Welles: Conversation at Oxford", *Sight & Sound* (Spring, 1960: Vol. 29, #2), page 82.

[2] Quoted by Pauline Kael in "Raising Kane" in *The Citizen Kane Book* (Boston: Little, Brown & Company, 1971), page 3.

> Why, despite the accuracy and complexity of its characterizations and the finesse of its effects, is CITIZEN KANE emotionally unengaging?
>
> What in particular sapped the *humanity* from CITIZEN KANE and precipitated Welles's transition to the design of THE MAGNIFICENT AMBERSONS?

A great film may influence films and filmmakers in one of three ways: positively, by encompassing innovations of such exemplary fruitfulness that they enter forthwith the mainstream of film design (THE BIRTH OF A NATION or UNDERWORLD, for example); negatively, by encompassing innovations of such exemplary unfruitfulness that they remain thereafter of only passing interest to film designers (NIGHT MAIL or THE LADY IN THE LAKE, for example); or both, by encompassing innovations both manifestly fruitful and unfruitful.

Films are rarely great for most are diffuse and hence unenlightening. CITIZEN KANE, however, is a film of the third kind – a movie both positively and negatively influential – and we shall misunderstand its influence and its lessons for us if we pretend otherwise.

Let us therefore list the tactical innovations of CITIZEN KANE, distinguishing the visual from the aural and the positive from the negative. By doing so we shall delimit the effects of radio drama on cinematical design.

The Tactical Innovations of CITIZEN KANE

Positive – Visual:

What visual innovations of CITIZEN KANE were productive, provoking emulation? Principally two.

1. The first – the sustained use of deep focus photography enabling viewers to see sharply all objects from foreground to background – had been attempted earlier by Eisenstein, Renoir and others. Success, however, awaited the invention of short focal length lenses with wide apertures, the development of fast panchromatic film stocks, the availability of sources of extraordinary amounts of artificial light and the imagination of Welles and Gregg Toland, his cameraman ("the greatest of us all" in Hal Mohr's estimation[3]), in forecasting its dramatic function.

[3] In the estimation of Hal Mohr, eminent cinematographer on THE JAZZ SINGER, THE WEDDING MARCH, A MIDSUMMER NIGHT'S DREAM, PHANTOM OF THE OPERA and THE FOUR

Deep focus had ancillary effects. Since action was now unrestricted to a single focused plane, events could be simultaneously recorded and contrasted at varying distances from the camera. Longer takes became possible from both stationary and moving cameras with less dramatic inelegance, and since follow-focus was unnecessary, longer takes permitted flexible camera movements (roving tracking shots, for example).

Deep focus, however, required visual content in breadth, depth and height for the wide angle lenses not only saw more than usual but registered it in focus. Sets, in particular, required ceilings, frustrating lighting directors constrained already to bring unaccustomed numbers of lights to bear on a scene. With the ceilings in place, however, low angle interior shots became possible for the first time, and Welles used them to show detail and decor.[4]

2. The second – the sustained use of low key lighting ("Rembrandt lighting" in Griffith's phrase: a few highlights surrounded by shadow) – resulted in part from the requirements of deep focus in both a superficial and a deep sense.

Deep focus required extraordinary amounts of light to register the images. With more objects seen sharply within a shot, more objects had to be lit. The fewer objects one needed to light, however, the more economical the production. Low key light was therefore the tactic of choice if deep focus cinematography was to be economical.

But there was a deeper impulse toward low key lighting inherent in the use of deep focus. For as Bazin soon noted,[5] if several characters are simultaneously in focus with a shot, they may equally draw attention and hence become dramatically ubiquitous. Hence the event may lose its dramatic momentum and become ambiguous, for each viewer is free to attend to whatever strikes his fancy.

POSTER, among others, speaking as a panellist in October 1973 at the symposium at George Eastman House, Rochester, New York, on "The Coming of Sound to American Film".

[4] Unlike Eisenstein, he was sensitive enough on most occasions to avoid using them symbolically.

[5] See in particular "The Evolution of the Language of Cinema" in Andre Bazin, *What is Cinema?*, Vol. 1, translated by Hugh Grey (Berkeley: The University of California Press, 1967), pp. 23-40. Portions of the surrounding essays, "The Ontology of the Photographic Image", "The Myth of Total Cinema" and the "Virtues and Limitations of Montage" are also pertinent, as are passages within his book on the work of Welles, *Orson Welles: a Critical View* (New York, New York: Harper & Row, Publishers, 1978) [French edition by Les Editions du Cerf, 1972], translated by Jonathan Rosenbaum with a Foreword by François Truffaut and a 'Profile' by Jean Cocteau.

Bazin, I think, misunderstood Welles's response to this constraint, for Welles faced the question head-on:

> Since all events within a shot photographed with deep focus may equally draw attention, how is it possible to constrain the attention of viewers to a *single* sequence of events as required for dramatic effect?

Welles's solution was to use low key lighting at all times (except when mimicking the newsreels), illuminating at any moment only that part of the scene to which attention should be drawn. He chose not to avoid shadows and accent highlights, the normal Hollywood lighting technique of the 1930s, but rather to light for the highlights and forget the unessentials. If, for example, the face of a character ought to be shadowed for dramatic effect, then shadowed it was, convention to the contrary notwithstanding.

The result was montage without cutting. Although the shots contained masses of detail, only significant details drew attention while the remainder registered through peripheral vision alone. Welles, contra Bazin, then sequenced the film around these points of attention maintaining momentum and forestalling ambiguity.

Positive – Aural:

What aural innovations of CITIZEN KANE were productive, provoking emulation? Again, principally two.

1. The first was an increase in the scope and accuracy of the sound effects contributing to the spatial sense of the scene (falling rain, for example, echoes in large rooms and the distancing effects of vanishing or approaching footsteps and voices).

2. The second was an increase in the effectiveness of the dialogue spoken by the characters compared to the 'theatrical norm prevalent in Hollywood filmmaking. Distinct voices were keyed to distinct characters (opposing characters? opposing voices!), and the naturalness of the spoken lines was augmented by using ungrammatical constructions, overlapping conversations and natural interruptions.

Let me now list the tactical innovations of CITIZEN KANE that were negative and have remained unemulated. All are techniques for switching from one scene to another with a time transition in between.

Negative – Visual:

What visual innovations of CITIZEN KANE were unproductive, remaining unemulated? Principally two.

1. Almost every technique used by Welles to indicate the passage of time between scenes proved to be dramatically counterproductive (the swish pans between the scenes of Kane and Emily at breakfast, for example, or the changing circulation figures for the *Inquirer*, the costume changes at the opera and the jigsaw puzzle montage).

The problem was not that the transitions were unclear, for every viewer could tell that time had passed and by how much, but rather that they were perceptually unnatural and hence emotionally empty; for, as Pudovkin would have predicted, a sustained emotional response can occur only when information is processed in perceptually natural ways. Emotions are felt responses occasioned by thwarted habits – physiological reactions to unconscious expectations gone awry, however pleasantly or unpleasantly. To evoke an emotional response other than that of naked novelty, therefore, requires the presence and prior assimilation of a context of natural stimuli naturally ordered, for only to natural stimuli naturally ordered do we possess deeply conditioned nonconscious patterns of response.

The visual time transitions of CITIZEN KANE, being perceptually unnatural, were perceivably only as symbols interrupting a non-symbolical context of events and hence, however clear their meaning, vitiated the possibility of sustained emotional response. Welles, sensing this upon later reflection, seldom again put them to use, and observant filmmakers have followed his example. Straight cuts (or, at worst, fade out – fade ins) now link scenes directly to one another with no disengaging clutter in between.

2. The second visual technique that was unproductive is the lapse dissolve in which a new scene is gradually introduced and superposed over the old, and the superposition is maintained momentarily before the old scene is faded out. Welles used the lapse dissolve to relate his character-narrators to the scenes following that they were supposed recalling (to place them visually 'on top' of the succeeding scene, as it were).

The lapse dissolve, like the devices of the first type, is a technique of time transition. It proved to be equally unnatural in dramatic contexts and hence equivalently emotionally stultifying, for we are unaccustomed to perceiving visual mixes and have no responses conditioned to them from which deep emotional effects could be induced.[6] Welles,

[6] See also page 11 below.

again sensing unfruitfulness later, avoided the lapse dissolve in his future designs, and most filmmakers have followed suit.

Negative – Aural:

What aural innovations of CITIZEN KANE were unproductive, remaining unemulated? Again, principally two (each again a technique of time transition).

1. The first used spoken lines of character-narrators as triggers to flashbacks. The idea, again, was clear, but clarity came at the expense of naturalness and emotional continuity. For at one moment we are watching a scene of a person reminiscing and are responding naturally to the events of the scene that we see and hear, while at the next moment we are shown a scene that the narrator is recalling, as if 'in his mind's eye', that we could not, in context, see or hear. The problem comes not from flashing back to another scene but rather from establishing the flashback as a memory image – an image in the mind of a character – that destroys the perceptual integrity of the establishing scene.

The problem lies embedded in the large strategical hassle of using perceived narrators in films. For the moment, however, we need only note that Welles was progressively to forsake the technique as have most other filmmakers.

2. The second unproductive effect occurred whenever a transition was made visually from one scene to another while maintaining an uninterrupted synchronous sound continuity begun in the first scene (as in the Thatcher montage or the election speeches when the transitions occur in the middle of a sentence begun in one scene and completed in the next; or when we cut in mid-phrase from Susan at her singing lesson to the opening of the opera; or when we cut between two scenes at different times of Kane doing tricks for Susan while maintaining a single uninterrupted conversation.)

As with the swish pans at the breakfast table, the transitions entail unnatural perceptions. The idea again is clear, but the effect, at best, is clever, and perceived cleverness is an undramatic intrusion sapping emotional continuity. Never again was Welles to use the technique, and sensitive film designers, shunning inelegance, have avoided it as well.

We thus have four listings of the tactical innovations of CITIZEN KANE (see Appendix to this essay). The historical question becomes:

> What led Welles to the positive and negative innovations of CITIZEN KANE? (That is, what were their distinguishing roots?)

The answer, as we shall see, was a susceptibility to the traditions of radio drama.

The Legacy of Radio Drama

Welles had been trained in the theatre, but his genius had found its focus in the new art of radio drama. He was brought to Hollywood to emulate the success of his War of the Worlds broadcast that had terrified the nation.

Radio drama, however, unlike the cinema, is a quintessential art of the ear deriving its effects solely from the use of voices, music and natural or unnatural sounds. What, then, are the virtues and limitation of the way we hear and hence of the design of radio drama?

Our ears, unlike our eyes, have no lenses. They register the presence of all sounding events in our environment simultaneously and indiscriminately. Unsurprisingly, therefore, our aural sense of the world at any moment is of a continuous spatial environment unbounded in any dimension.

Our aural sense of the world, however, is also uncacophonous, for at any moment we are attending (usually) to only one sounding event while registering all others peripherally. Since our ears have no lenses, they are unable to focus on particular sounds to the exclusion of others simultaneously registered. Why, then, is the aural world uncacophonous? Or, more precisely,

> How is it possible for us to *attend* to a single sounding event within our environment having no lenses with which to *focus* upon it?

This perplexing problem, known affectionately as 'the cocktail party problem' in American acoustical science (How are we able to distinguish one drunken voice from another in a crowded room?), has only recently been solved. Our ears are frequency analysers. When sound waves are received by one of our ears from a sounding object, they are (usually) out of phase with the waves received by the other ear, and both sets of waves are out of phase with the waves received from other sounding objects in different spatial locations. It is this difference in phase of the waves received by each of our ears from spatially disparate objects that permits the brain to attend to one object rather than another even though our ears have no lenses with which to focus on it.

The higher regions of our aural perceptual system, in short, encompass a complex phase filter. For the phase filter to work, however, the waves striking our ears from different objects must come from different spatial directions! Otherwise there would be no phase difference between the waves to be filtered and analysed.

But therein lies the source of the fundamental problem of radio technique, for all sound waves emanating from a single radio speaker arrive at our ears from a common spatial direction in phase! Were one, therefore, to playback through a single speaker with absolute fidelity a recording taken in the middle of a room at a cocktail party, one would be unable to select from the mass of sounds a single sound to which to attend, even though, had one been standing in the middle of the room at the cocktail party registering the same mass of sounds, one could have done so with ease.

Designers of radio drama, therefore, faced a tricky problem. Listeners must attend to single events within a natural aural environment to become emotionally engaged. But an accurate phased reproduction of an aural event within its environment was impossible. Aural scenes, therefore, had to be constructed, not reproduced.

By 1941 designers of radio drama had learned to mimic our natural aural experience of a scene by

1. sustaining a sense of a continuous sequence of single events to which attention could be drawn by presenting at attentive levels only those events and eliminating (or attenuating) all others;

2. sustaining a sense of a continuous and unbounded spatial environment by maintaining a constant, unobtrusive background presence (room tone, relevant noises, etc.);

3. sustaining a sense of an accurate spatial distance between sounding objects by varying their relative loudness as they advanced or receded from the foreground; and

4. sustaining a sense of the immediate recognisability and unquestioned credibility of characters by using distinct voices for different characters and employing natural (nontheatrical) patterns of speech and conversation.

The sources of Welles's positive innovations ought now to be apparent. Welles came to Hollywood accustomed to conceiving of a scene as a sequence of single spatial events of unquestioned natural credibility existing within an unbounded spatial environment simultaneously present but unattended. When he attempted in CITIZEN KANE to

recreate the natural sense of the scene to which he was accustomed in radio, the results were cinematically ingenious and fruitful:

> By using deep focus and low key lighting (and their concomitants), Welles achieved, through natural visual techniques, the effect of unbounded spatiality with bounded attentiveness to which he was accustomed; and
>
> By transferring to the soundtrack the detailed sound effects to which he was accustomed (the effects of presence and distance, the distinctive voices of characters, the natural patterns of speech), Welles accentuated the natural spatial credibility of the events visually perceived.

What, then, were the sources of Welles's negative innovations in CITIZEN KANE – those effects that he later recognized as unproductive and that have remained unemulated?

Our ears are marvellous tools for ascertaining the presence and distance of identified sounding objects. Unfortunately, lacking lenses, they are astonishing poor at ascertaining the spatial directions from which sounds come.[7] As Aristotle knew, therefore, and Leonardo da Vinci reaffirmed, our eyes are our chief tools for ascertaining the spatial identity of objects and events (that is, the spatial map of our environment).

Furthermore, most of the cues that the environment provides for the chronological identity of objects and events are visual. Our ears are awesome distinguishers of the subtle rhythms and pacings with which our aural environment unfolds itself, but few of those temporal signals are historically precise. We hear the wind in the trees and then the rain, but the succession sounds too much like the wind and the rain of two weeks ago to be of much chronological use. Voices aside, most of our information about the chronological progression of the world comes through our eyes.

Radio dramatists, therefore, faced a second tricky problem. Listeners must be oriented to the spatial and chronological map of their environment to become emotionally engaged. As suggested above, this must usually be done by describing the dramatic environment to them with a human voice, for all other aural tools are insufficiently precise. But describing an environment is an unnatural use of the human voice in most situations, for we are accustomed to acquiring such information through our eyes, not our ears! (When walking with a friend, conversing, one is unaccustomed to saying or hearing, for example, 'Over there is a yellow automobile, and there a brick building, and

[7] Experimental subjects frequently misconstrue directions of impinging sound sources by 180 degrees!

over here ... ' One may comment upon the environment seen or express evaluations of it, and listen as others do so, but one seldom if ever describes it or hears it described. Verbal descriptions customarily inform us of past or future events, not events present to us.)

Radio dramatists were therefore caught on the horns of a dilemma:

> A radio drama ceases to develop if its environments remain undescribed, for listeners are then unable to identify the spatial and chronological setting of the action; but if the *characters* describe their environment, they then lose credibility as characters, and (again) the drama ceases to develop.

To avoid the dilemma (to protect the credibility of the characters, that is, while permitting listeners to identify the dramatic environment), radio dramatists developed a tool with precedents in the novel and theatre but of genius in the new medium alone: the narrator.

By using a voice unlocated in the scene, a dramatist could identify for listeners the spatial and chronological environment of events with neither fragmentation nor loss of character credibility. Time transitions, in particular, could be made with simple elegance by shifting the tense or grammatical voice of a narrator's remarks: flashbacks and flashforwards could be effected with precision and consummate ease.

When Welles came to Hollywood, he was accustomed to achieving time transitions flexibly and accurately by means of narrators. Unsurprisingly, therefore, he (and Mankiewicz, his co-scriptwriter) designed CITIZEN KANE as a sequence of scenes encompassing major and frequent time transitions. The story of Kane's life is presented chronologically (with minor overlaps) in four major flashbacks each containing multiple scenes having major flashforward time transitions between them. The film ends where it began – in Kane's mansion near the time of his death.

Given the strategy, the tactical problem was clear:

> What techniques could be used to effect the time transitions while maintaining dramatic continuity and momentum?

Accustomed to the sounds of radio but not the sights of cinema, Welles – with astonishing ingenuity – construed the sights as sounds and subjected them to sound techniques! The results, however, were unproductive in two ways.

Welles did not wish to forgo the flexibility and ease with which narrators can effect time transitions. He did however wish to utilize the immediate identification of characters that visuals provide. He decided, therefore, to use characters as quasi-narrators to introduce and identify the flashbacks – not one character, but four!

Welles recognized that the only way to introduce a flashback credibly through a character narrator in a scene is to treat it as if it were a memory image. Unfortunately, the introduction of a purported memory image destroys the integrity of the scene, for memory images are unperceivable by observers, unlike the face and voice of the character introducing them.[8]

The second failure arose again from a lack of understanding of the non-equivalence of vision and audition. Our ears, having no lenses, register complex mixes of sound as multiple sound events superposed on one another. We are therefore accustomed to perceiving aural mixes (the sounds of a cocktail party, for example) and sorting them out. Our eyes, on the other hand, having lenses, focus on objects one at a time. Visual mixes or superpositions are not part of our everyday visual environment.[9] We are unaccustomed to perceiving them and hence can have no rooted emotional responses to them when such perceptions are cinematically contrived.

Welles was accustomed in radio to using aural mixes and superpositions (cross fades, for example, and sound overs) to effect time transitions with natural dramatic results, for aural mixes are part of our everyday aural environment, and we have conditioned nonconscious responses to them. When, however, he attempted in CITIZEN KANE to use visual mixes and superpositions to effect time transitions or to unify disparate events, the results were unnatural and hence emotionally vacuous. When he attempted similarly to bridge disparate visual scenes with continuous synchronous sound, the results were again unnatural and dramatically unproductive, for we have no conditioned responses to such perceptual effects either.

The sources of Welles's negative innovations ought now to be apparent. Welles came to Hollywood accustomed to the flexibilities of audition. When, however, he attempted in CITIZEN KANE to treat visuals as sounds (and hence flashbacks as memory images), and

---

[8] Perhaps disquieted by the effect of the multiple-character narrators of CITIZEN KANE, Welles tried a voice-over non-character narrator in THE MAGNIFICENT AMBERSONS and then a voice-over single character narrator in THE LADY FROM SHANGHAI. Finally, recognizing the root of the problem to be his vestigial (literary) insistence on major and frequent time transitions rather than the narration *per se* (Bresson's voice-over narrations, and Welles's own in THE MAGNIFICENT AMBERSONS are, after all, elegantly effective), Welles eschewed all three in TOUCH OF EVIL with brilliant results.

[9] See page 5 above. Even colours are never seen *as mixes*.

subject them to sound techniques, the results were uniformly clever but unproductive. For the events became symbols and sapped the emotional development of the drama.

The Legacy of CITIZEN KANE

CITIZEN KANE, despite its brilliance, is emotionally unengaging for its major and frequent time transitions, derived from radio drama, compelled Welles to use tactics that were perceptually unnatural. Had they been less unnatural, the film would be less brilliant. As it is, however, CITIZEN KANE's may (de)vices cancel its virtues: its surface engages the mind while disengaging the heart. Welles recognized this more quickly than anyone else, and when, one year later, THE MAGNIFICENT AMBERSONS appeared, the compulsions of radio drama were (largely) gone forever.

We can learn much from CITIZEN KANE about the virtues and limitations of radio and the cinema, but not by shutting our eyes and ears, closing our minds and shouting 'masterpiece'. We must instead construe it as Welles did – as a "catalogue of effects" – most of them begotten by another art, radio drama, and half of them stillborn.

Only by following Welles example shall we come to understand how great an achievement CITIZEN KANE really was.

Error! Bookmark not defined.**Appendix**

The Tactical Innovations of CITIZEN KANE

Positive – Visual:

1. The use of *deep focus* permitting simultaneous actions in diverse planes, flexible camera movements and longer takes and the use of full sets with ceilings, permitting low angle interior shots.

2. The sustained use of low key ("Rembrandt") lighting permitting 'montage without cutting'.

Positive – Aural:

1. The use of abundant and accurate sound effects to contribute to the spatial sense of the scene; and

2. The use of unprecedentedly effective dialogue achieved through

   (a.) distinct voices keyed to distinct characters and

   (b.) natural constructions of speech (ungrammatical sentences, for example, and overlapping and interrupted conversations).

Negative – Visual:

1. The use of visual devices to indicate time transitions, especially

2. The use of lapse dissolves.

Negative – Aural:

1. The use of spoken lines of character-narrators as transitions to flashbacks; and

2. The use of uninterrupted synchronous sound continuities to bridge visual time transitions.