Viewpoint, misdirection, and sound design in film: *The Conversation*

Vera Tobin [*]

Case Western Reserve University, USA

Abstract

Stories can and often do build surprises by encouraging audiences to attribute certain assertions, presuppositions, and evaluations to an ''objective'' or base-level perspective, only to reveal later on that these elements should be attributed only to the mistaken or deceptive viewpoint of a particular character. This paper presents a comparison of sound design and viewpoint phenomena in Francis Ford Coppola's film *The Conversation* (1974) with similar narrative twists in prose and with other perspective shifts in film. It shows how viewpoint blends, shifts, and distinctions between the ''viewpointed'' and ''non-viewpointed'' status of elements in the visual and auditory stream in film can work together to create this kind of re-evaluation surprise, and discusses how these are and are not analogous to similar effects in prose.
© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Both linguistics and literary studies have a long tradition of comparing resources for reporting speech, thought, and perception across different media, genres, modalities, and settings. These comparisons often take a special interest in phenomena where the exact status of the representation involved is unstable, difficult to classify, or otherwise somehow ''mixed.''

Free indirect discourse, for instance, is relatively rare in conversation—rare enough that for some time, many scholars (e.g. Lips, 1926; Banfield, 1982; von Roncador, 1988) thought it didn't exist outside written discourse at all. That has turned out not to be the case. As Fludernik (1993:83–90, inter alia), for one, has shown at length, free indirect discourse does indeed appear in oral narratives, as well as in journalism, literary representations of dialog for stage, verse, and elsewhere. Meanwhile, speakers often produce gestures that express a different viewpoint than the one(s) expressed in their speech; what's more, they also can produce *dual viewpoint* gestures (McNeill, 1992; Parrill, 2009) that themselves express multiple perspectives on a scene or event at the same time. Signed languages, too, offer a variety of resources for presenting multiple perspectives in a single utterance, as seen for instance in body partitioning (Dudis, 2004) and fused perspective constructions (Perniss and Özyürek, 2008). Generally, the more we look, and the wider the range of data we consider, the more clear it becomes that constructions evoking multiple viewpoints at once are ubiquitous across genres and modalities (Dancygier et al., 2016).

Additionally, mixed and ambiguous viewpoint constructions have proved an extremely rich and productive resource for storytellers who want to produce certain kinds of surprises. This paper takes up a particularly narrative viewpoint

---

 * Correspondence to: Case Western Reserve University, 617G Crawford Hall, Cleveland, OH 44106-7063, USA.
   E-mail address: vera.tobin@case.edu.

phenomenon: story twists that hinge on the re-analysis of what, or whose, viewpoint should be associated with a given element in the narrative. Stories can and often do (see, e.g., Emmott and Alexander, 2010, 2014; Pallarés-García, 2012, 2014; Tobin, 2009, 2014) construct surprises by encouraging audiences to attribute certain assertions, presuppositions, and evaluations to an ''objective'' or base-level perspective, only to reveal later on that these elements should be attributed only to the mistaken or deceptive viewpoint of a particular character. This happens commonly in written narratives, but also in drama and film.

We will look in detail at one distinctive instance of this viewpoint gambit as enacted in film: the set-up and climactic twist of Francis Ford Coppola's 1974 *The Conversation*. This film is noteworthy for its virtuoso management of multiple viewpoints across and between sound editing and the visual stream. It also famously includes one specially daring bit of editorial sleight of hand, which will be the primary focus of the present analysis.

The Conversation tells the story of a surveillance expert working on a particularly difficult piece of audio reconstruction. For the first third of the film, the crucial snippet is indecipherable. We see and hear the process of juxtaposing and combining the information from different tapes, until their words emerge from the muddle. Later, in the film's final moments, the expert realizes that he was mistaken in his interpretation, and we hear (as he does) a new version of the recording.

In this last ''replaying,'' one section of the original, oft-repeated recording is not just altered but replaced. The new version places stress on different words than the original, producing different implications about the speaker's intentions. How the film manages that switch, the risks it poses to cooperativity, and its closest neighbors in prose and cinema, are the focus of the present analysis. What we will see is that while the mixed/ambiguous status and its discourse functions have many important parallels to free indirect discourse, the material particulars of how the filmic versions are constructed are not negligible.

In looking in detail at these mixed viewpoint representations in *The Conversation*, we can learn more about:

- Medium-specific strategies for manipulating reader attention and depth of processing (Sanford and Emmott, 2012) through foregrounding and backgrounding, and how these intersect with viewpoint in different modalities;
- Mixed viewpoint representations involving both the visual and the auditory stream: in film, these can be decoupled from one another to create a number of striking effects;
- The importance of embodied viewpoint as a resource for building narrative and discourse coherence; and
- Issues of inconsistency, unreliability, and uncooperative narration (Grice, 1975; Kukkonen, 2013) that arise around these viewpoint manipulations.

2. *The Conversation*: plot summary

A San Francisco business executive has hired Harry Caul (Gene Hackman) to spy on the executive's wife. As events unfold, Caul becomes both increasingly obsessed with uncovering the truth about his client and increasingly paranoid about his own vulnerability to surveillance. Caul's inner turmoil and the external mystery both play out through his compulsive attentions to the product of his work: the tapes of the eponymous conversation.

The film opens with a complex set piece in which Caul and his team use an array of microphones to record a pair of lovers—Ann (Cindy Williams) and Mark (Frederic Forrest)—as they walk slowly around San Francisco's busy, noisy Union Square. Later, as Caul compiles the recordings, elements of this scene are replayed for the audience many times. We come to know the crucial parts very well, and one moment best of all: a point at which the words are at first almost completely drowned out by music and other noise. Caul plays and adjusts and replays and refines the recording of this moment (as we look and listen on) until it is clear: Mark is saying ''He'd *kill* us if he got the chance.'' Caul concludes that his client, Ann's husband, has commissioned the surveillance to get proof that Mark and Ann are having an affair, and that their lives are in danger as a result.

Caul, already consumed with guilt over a past wiretap job that ended in murder, becomes desperate to protect the lovers. He refuses to turn the tapes over to his employer and goes on his own mission to try, somehow, to save Mark and Ann. But ultimately it turns out that the conspiracy was not a plot by Ann's husband, ''the Director,'' to kill the couple. Instead it was their plan to murder *him*. As Caul realizes, too late, what has happened, we again hear Mark's taped line. In this final iteration, however, it's changed: ''He'd kill *us* if he got the chance.'' This is a different recording from the one we've heard before—not a cleaned-up version of the same clip, but an entirely new one, with a different line reading.

The film's editor and sound designer, Walter Murch, explains in an interview (Koppelman, 2005:38–39) how it was done: during production, Forrest and Williams recorded additional takes of their lines from the Union Square scene in a quiet room, in case anything was unusably unclear in the original footage. In that session, Forrest flubbed one take. He delivered the line not as directed, with the prosody that suggested the couple was in danger, but with the emphasis on ''us''. At the time, Murch set the material aside as a mistake. But later, when audiences for test screenings were having trouble understanding the film's plot, he pulled the clip from his archive. This version of the line perfectly encapsulated the interpretation that Caul has missed; as Murch describes it, ''the implied conclusion, 'Therefore we have to kill him.'''

Using two different performances of the crucial line was ''a risky idea,'' Murch acknowledges, because it threatened ''one of the fundamental premises of the film, which is that the conversation itself remains the same, but your interpretation of it changes.'' But, interestingly, audiences found this version of the film much clearer, more satisfying, and more comprehensible than earlier edits.

3. Inconsistency, viewpoint, and threats to cooperativity

There is a puzzle here: Murch and Coppola (DePalma, 1974) both point to this swap-over as a vital ingredient in helping the film's narrative hang together for viewers. But trading one version of an important sequence for another, while still treating the two as repeated presentations of the same event, runs a serious risk of fatally violating the rules of cooperative discourse (see, e.g. Pratt, 1977) for narrative fictions.

Indeed, there is an illustrative, and notorious, close parallel in the history of Hollywood film. In the first half of the twentieth century, Hollywood movie studios produced a steady supply of weekly adventure serials: action-packed, cliffhanger-laden tales doled out one Saturday matinee episode at a time. These ''chapter plays'' could be Westerns, science fiction, romance, war stories, mysteries, the adventures of a gentleman thief, or nearly any other sort of story. Whatever the setting, the structure and pacing were typically very much the same. Each chapter but the last ended with the protagonist in the middle of a suspenseful predicament. The next episode took up the events, not where they left off, but a beat or two earlier, reprising the final exciting scenes of the episode before. The second time around, though, the sequence would continue, revealing how the hero managed to escape what looked last time to be his certain doom. Sometimes these cliffhanger resolutions took advantage of the gap between viewings to revise the original depiction of events—for example, clearly showing characters still inside a boat as it hits a rocky shoal (in ''King of the Texas Rangers,'' episode 3), only to show them jumping out well before impact on ''replay'' the next time around. Harmon and Glut (1972) report that some studios were especially notorious for indulging in these shenanigans, which they characterize as ''always booed'' (p. 362).

What makes the editorial shenanigans of *The Conversation* different from the chapter story cheats? In both, there is a sequence that is recorded, and presented, one way. Then, later, the ''same'' sequence is ''replayed''—only this time, it's not the same at all! The key difference is that, unlike the misbehaving movie serials, *The Conversation* offers a way of understanding the non-replication while preserving the conceit that ''the conversation itself remains the same.'' The requirement is that we must understand all the clips as having been, at least in part, instances of represented thought and perception—dramatizations of Caul's own mental soundscape.

In the case of *The Conversation*, the subjective framing is retroactive. The audience is not encouraged to interpret the sound clips as representations of character viewpoint until the critical moment. How is this managed? Strategic foregrounding and backgrounding, or ''burying,'' as Emmott and Alexander (2010, 2014) call it, is crucial for leading the audience to accept certain details as unviewpointed but also accept later that—surprise!—they were representations of a blended character/narrator viewpoint all along. A narrative can avoid seeming illicitly inconsistent if viewpoint cues are subtle and backgrounded at first, then foregrounded later on, in concert with narrative shifts. Prose offers many resources for doing this sort of thing, including the use of narrated perception and free indirect speech and thought.[1] *The Conversation* uses a number of medium-specific methods, some closely parallel to FID and others less so, to achieve a similar effect. The particulars of these strategies, and of the division of labor between the visual and the auditory presentation of information, also reveal the tremendous value that embodied viewpoint offers for both manipulating audience attention and creating discourse coherence.

4. Resources unique to the medium: cross-modal shell games and the conventions of continuity

One very important strategy that *The Conversation* employs to finesse the viewpoint status of elements capitalizes on components of the cinematic editing style known as *continuity editing*. For the film's narrative to succeed, audiences must be aware of the limitations of what characters can know on the basis of their perceptions—indeed, those limitations are actively thematized in the film's focus on surveillance and suspicion. At the same time, the degree to which those perceptions may be not only limited, but also *mistaken*, must be downplayed until the crucial moment. The structural role that viewpoint cues play in the conventions of Hollywood editing is important in making this balancing act possible.

The depiction of multiple viewpoints within a narrative is often an expressive goal in itself. But visible signs of character viewpoint can also serve as resources to be exploited for the sake of achieving other discourse goals, most notably goals

---

[1] Sanford and Emmott (2012) offer a thorough survey of linguistic resources for managing readers' depth of processing, both positively (through foregrounding) and negatively (backgrounding).

Fig. 1. The vantage from above.

of continuity and coherence. This fact allows *The Conversation* to associate material with the perspective of a specific character or characters without foregrounding that association or even, as we'll see, fully committing to it.

The conventions of classic Hollywood-style continuity editing, developed over the course of the twentieth century (Bordwell et al., 1985), comprise a suite of standard methods and principles for keeping viewers oriented in time and space across these transitions, producing the sense of a unified, "continuous" narrative progression. The object of this style is to lead viewers, without much or any conscious effort, to knit disparate elements together across cuts into what feels like a smooth and coherent whole.[2] Importantly, many of the conventions of continuity editing work by tapping into viewers' attunement to physical signs of viewpoint that manifest in performers' bodies, in keeping with Sweetser's (2012:13) observation that "there is no more powerful icon for a bodily viewpoint than an actual body with an actual inherent viewpoint."

For example, the opening sequence of *The Conversation* presents a physically complex scene, in which the viewer is asked to keep track of physical relationships across many different players moving in different directions around a crowded area. It opens with a slow zoom from a crane shot above Union Square, which establishes the layout of the scene and, as the zoom moves in on a smaller portion of the action, focuses our attention on specific players. As this is happening, the street sounds, too, get louder, and are punctuated by strange sounds that eventually turn out to be artifacts of the recording apparatus Caul's team is using. This zooming camera-eye cannot be depicting the visual viewpoint of any of the people walking around at street level. This angle on events might, of course, be purely narratorial: not tied to the viewpoint of any character within the world of the film. However, we soon learn to understand that it is motivated by, and associated with, the presence of a watching and *listening* vantage high above the square, through the sequence of cuts shown in Fig. 1.

At just over three minutes into the film, the zoom comes to rest—still from above, rather than on a level with the pedestrians' line of sight—on Harry Caul's trench-coated back (shot 1–1). We then at last cut away from this shot for the

---

[2] Of course, different effects can be produced, either deliberately or by accident. So-called "dynamic" editing techniques, in which cuts juxtapose elements or sequences rather than placing them in service of the feeling of continuity, diverge from the constraints of the continuity system for dramatic and affective intensity (Rosenberg, 2011). Montage sequences are an example of dynamic editing.

first time: to a rooftop sign atop a building nearby (1–2). A human figure is visible, crouched below the sign. As we continue to hear the (glitchy) stream of talk, music, and ambient noise of the square below, the film cuts to a closer view on the figure (1–3). He is recognizable as the same man shown in shot 1–2, thanks to the fact that his body is still in the same pose, though seen from a new angle. He is looking off-screen, through the viewfinder of a large directional microphone, and is wearing headphones. A close-up from the same angle (1–4) makes the details of his gaze off-screen even clearer. Finally, we see the street scene again (1–5), from a similar angle to that of shot 1–1, but now focused on a man and woman (Mark and Ann) and with new visible elements suggesting that we are seeing exactly what the rooftop listener sees: most notably, the crosshairs of the viewfinder, but also (1–6) the inclusion of obstructions to his/our line of sight as the camera tracks their movements.

Eyeline match shots like these, a staple of continuity editing, cue viewers to physical and attentional connections between elements within a scene. These cues are a matter of established convention, but they also, of course, rely on viewers' tendency to notice gaze direction, and their inclination to draw inferences based on assumptions of relevance. In an eyeline match edit, one shot presents a character looking at something outside the frame of the shot, preceded or followed by a shot presenting the seen space, sometimes but not necessarily from the exact vantage of the viewing character. In continuity editing, the composition of many shots is thus structured with close reference to character viewpoint, but what is shown on screen need not, and typically does not, have the status of a direct representation of the character's perceptions.

Eyeline matches and other sorts of reaction shots *can* set up fully depictive point-of-view sequences, as seen in the crosshair view described above. But the relationship between characters' viewpoint and the depiction of that which we understand them to be looking at can also be partial and ambiguous, as with the opening zoom shot. The physical arrangement of characters' lines of sight commonly serves to provide an organizing logic and coherence to the angle and sequence of shots in a scene, without suggesting that any of those shots should be taken as depicting character viewpoint in a strict sense.

Indeed, this state of affairs is the typical one. A standard Hollywood approach to staging dialog scenes, for instance, is to show one character looking at another character, usually off-screen, cross-cut with shots of the other character, looking back at the first (shot/reverse shot). The camera is typically to one side of the rear-facing character's original position. In keeping with this arrangement, the forward-facing character typically does not look directly into the camera. As such, the characters' gaze provides orientation for viewers between cuts without necessarily generating the implication that any shot constitutes a strict representation of the point of view of any of the characters in the scene. As Parrill et al. (2016) have shown, viewers who have no trouble following the transitions of shot/reverse shot editing nonetheless tend not to reproduce these viewpoint shifts in their own speech, gesture, or sign when asked to recount scenes that include it. Because these editing conventions are so pervasive, audiences are trained to track these viewpoint cues without foregrounding them or necessarily attributing their counterpart shots to character viewpoint. Instead, they seamlessly integrate them into an understanding of the scene as a whole, forgetting that many shots and sounds have been set up with reference to different character viewpoints.

Associating the opening zoom with the perspective of the microphone operator stationed on the rooftop nearby thus first and foremost works to help us make sense of the physical and dramatic relationships between the eavesdroppers and the events on the ground. It serves the exigencies of continuity and of establishing the *mise en scène*. This association is also necessary if we are to make sense of those strange disruptions in the soundtrack. Or, to think of it another way, the mysterious noise in the auditory channel is part of what prompts us to connect the establishing shot to the rooftop perspective. By virtue of their oddness, the noises draw attention to the experience of hearing, and suggest that these sounds are not unmediated. When the microphone operator appears, the auditory mystery is resolved.

This sequence establishes a pattern that continues through the film: when the camera view is roughly *aligned with*, rather than necessarily *taking the exact position of*, a viewing/viewed character's subjective vantage, what we hear may later turn out to have been more tightly connected to a limited perspective. In other words, where the camera lines up with character viewpoint for what at first seem to be primarily reasons of physical coherence, the alignment is latently available to be re-interpreted, via an auditory prompt, as a direct representation of character viewpoint. This pattern, and the relatively loose alignment of mixed viewpoint representations in the visual and the auditory streams, lays the groundwork for the progression of variations on the garbled audio culminating in the final twist. Because the pattern is in line with common conventions of editing for continuity, the discrepancies are not obtrusive. The frequency and ambiguity of viewpoint cues in these scenes are easily accommodated because the cues have multiple reasons to be there.

At other points, *The Conversation* plays even faster and looser with the relation between what is heard and the viewpoints we see on screen. For instance, midway through the Union Square stakeout, Caul retreats to a parked van with mirrored windows, where his assistant Stanley (John Cazale) sits, headphones on, operating an array of recorders. We hear Mark and Ann's conversation playing on while the camera tracks Caul walking across the street and entering the van. When the visuals cut to the van interior (5′14″), revealing Stanley and his equipment, the implication is that the current soundtrack—still the conversation from the square—maps onto whatever is being piped in through those headphones.

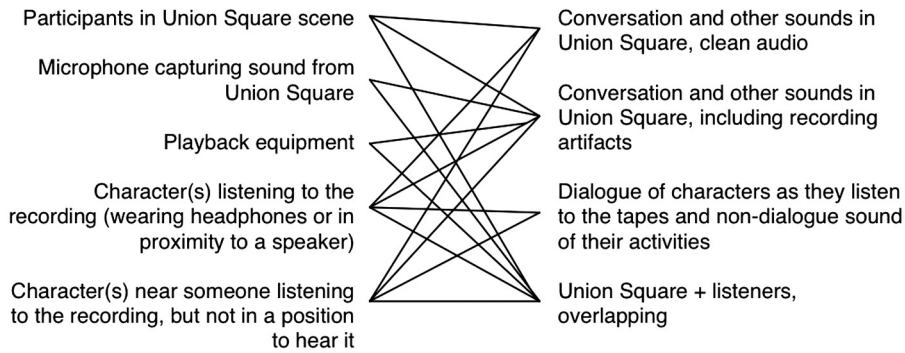|	
Participants in Union Square scene	Conversation and other sounds in Union Square, clean audio
Microphone capturing sound from Union Square	Conversation and other sounds in Union Square, including recording artifacts
Playback equipment	Dialogue of characters as they listen to the tapes and non-dialogue sound of their activities
Character(s) listening to the recording (wearing headphones or in proximity to a speaker)	
Character(s) near someone listening to the recording, but not in a position to hear it	Union Square + listeners, overlapping

Fig. 2. Cross-modal combinations in the van sequence.

That relationship established, the auditory stream then becomes more complicated. The surveillance recording plays on, but overlaid in the sound mix we now also get the dialog and ambient noises of the action inside the van. Caul and Stanley talk briefly about the rooftop operator. The camera cuts to him, still crouched under the sign. Then Caul (audible, not visible) asks, ''How about second position?'' prompting a quick pan (5′43″) to an open window with another man, operating another microphone. The sound quality of the conversation, still playing in the background (see section 5.2), declines significantly through a time-matched crossfade to what we can take to be the audio from the window microphone.

As the sound from ''second position'' continues playing, the visuals cut back to the van interior (5′46″): ''It's not so good,'' Stanley says, and makes an adjustment to the recording equipment, so that, in the words of the original screenplay (Coppola, 1972:7), ''for a moment, the conversation is doubled up on itself.'' Finally, we again see Mark and Ann, conversing (5′50″), but now matched to the inferior audio: full of gaps and distortions. Later (beginning at 6′03″), the sounds of Union Square drop out of the mix entirely for a short period, when Stan, still wearing the earphones, is distracted by a pair of pretty women looking at their reflections in the van's mirrored window. ''Pay attention to your recordings,'' Caul says, prompting a cut back to the square and the return of the surveillance audio to the soundtrack (6′28″).

This sequence establishes a number of important facts about the circumstances of the surveillance situation and the tapes that are being produced: there are multiple recordings being made simultaneously, from different angles; these recordings are not all equally ''good''; and they can be combined in different ways to approximate the most ideal rendition. It also sets up a number of different audiovisual combinations, as diagrammed in Fig. 2, all of which recur repeatedly during the portions of the film in which Caul edits and replays the surveillance tapes.

Throughout, sound freely moves between being only very loosely connected to character viewpoint—as when Caul is outside the van, or later (6′40″), when Stan hands the headphones to him without any corresponding break or modulation in what the audience hears, even during the transition when neither character is in a position to hear the playback—to being tightly bound to character viewpoint, as when the playback disappears from the soundtrack in accord with Stanley's failure to ''pay attention to [his] recordings.''

Cues regarding the viewpoint status, physical location, timing, or dramatic significance of what is *seen* can thus mingle with, hint at, or diverge from various aspects of the status of what is *heard.* This freedom to mix and match audible and visible counterparts along different dimensions, buoyed by the conventions of continuity editing and viewers' relative lack of focus on these elements of cinematography, makes the film's cross-modal viewpoint shell game possible.

5. Viewpoint management by other (multimodal) means

Now we will survey other viewpoint-relevant audiovisual methods that the film uses to downplay the significance of crucial information and set up the final twist. *The Conversation* exploits the affordances of its medium to deploy several complementary strategies for burying information (Emmott and Alexander, 2014) to produce a satisfying narrative surprise.

5.1. Foregrounding the source

Patterns of cross-cuts can encourage *selective focus* (Barton and Sanford, 1993) on specific elements of a scene. The underlying technique is backgrounding by way of foregrounding—drawing our attention to one thing in order to keep it away from something else. Here we will look at how *The Conversation* achieves this effect through literal foregrounding of the equipment used to adjust and replay the Union Square recording while the crucial conversation is heard. This maneuver exploits not only the audiovisual affordances of film, but also the multidimensionality of viewpoint itself.

There are two main sequences in which Mark and Ann's conversation is played, adjusted, and re-played to produce the first clear audio of the initial line reading ''He'd kill us if he got the chance'' (audible at 37′50″). In them, the number of reaction shots showing characters listening to the playback (27) is nearly perfectly balanced with the number of close-up shots of playback equipment (26): tapes whirling, the speaker mounted in a high corner of Harry Caul's work space, control dials and buttons of the mixer. In addition, these sequences include eight medium-range shots presenting both listeners and equipment together in the same frame, as well as establishing shots beforehand in which the tapes, speaker, and mixer feature prominently. Indeed, it is only quite late in the film (1.25′49″) that we ever hear any part of the conversation in a sequence during which no audio equipment is visible, the last scene before the night of what turns out to be the Director's murder.

By visually foregrounding the technology of recording and playback devices, *The Conversation* emphasizes physical limitations on character perceptions: the material conditions that may distort or obscure the sound. This foregrounding serves to de-emphasize the evaluative dimension of personal viewpoint, discouraging premature inferences about the fallibility of these perceptions along those lines.

Emmott and Alexander (2014:332) describe this burying technique as follows: ''Stress one specific aspect of the item so that another aspect (which will eventually be important for the solution) becomes less prominent.'' They note that character attention (say, to eyecatching color) can be a very useful pretext for this method. licensing the mention of attended-to qualities of an object while passing over others. Here, the technique works specifically to emphasize and de-emphasize different aspects of character viewpoint. Attention is directed toward what it is *possible* for a character to perceive given his location and the material conditions that obtain, on the one hand—such as line of sight, earshot, the affordances of the equipment—and away from his introspective *interpretation* of what he perceives.

5.2. Foreground and background within the auditory channel

The Conversation also makes use of sonic foregrounding and backgrounding within at least three different sorts of overlapping sound. These methods encourage selective focus and help to set up the final viewpoint twist, complementing the audiovisual methods described above.

5.2.1. Foregrounding of non-dialogic sounds

Several factors can affect the perceived prominence of non-dialogic sounds in a scene (Sonnenschein, 2011; Teki et al., 2011), including relative volume, novelty, whether or not they are keyed to events visible onscreen (Chion, 1994:69 notes that images can ''magnetize'' sound for viewers), and whether or not they co-occur with dialog (people preferentially attend to speech). It would be odd if the scenes where we see Caul working with the tapes had no sound effects depicting the noise of the equipment itself. Instead, these sounds, which can also be heard more quietly and subtly in other scenes, are not just present but very prominent in the mix: they are loud, they are sudden, they are synchronized with close-up images of the events they depict, and they are intercut with the dialog rather than underlying it. These methods accentuate focus on the spinning tapes and the mixer, enhancing the effect of the visual foregrounding in these scenes.

5.2.2. Overlapping dialog

There are several instances of overlapping dialog in which the here-and-now of the listening characters is interpolated onto the there-and-then of the Union Square scene. During the original recording session, for instance, we see and hear Ann and Mark notice Paul, a member of the surveillance team, who has been sitting on a bench with yet another microphone. As the sound and image from the square continue, Stan's voice suddenly intrudes on top (6′35″)—''Well, that's it for Paul. She spotted him''—yanking us, sonically, back to the van. A similar interruption punctuates Caul's second editing session. Well into yet another rewind-adjust-replay, Mark and Ann walk (in flashback) and link arms. ''Do you think we can do this?'' says Ann, and then: ''What a *stupid* conversation!'' exclaims Stan's voice (33′28″). ''Stan, please,'' Caul's voice entreats. ''I'm tryna work.''

These intrusions do several useful things. They prevent us from being too fully immersed in the Union Square scene, forcibly reminding us of the lovers' constrained, observed state. They also remind us of the presence of other witnesses to the recordings, de-emphasizing the fact that only Caul is present for *all* of our encounters with them. Finally, they can distract us from picking up on clues in the lovers' dialog itself. What ''this,'' we might have wondered, is Ann uncertain about doing? But we won't dwell on the question too long, because Stan's interruption has stolen focus.

5.2.3. Ambient sounds

In the idiom of contemporary film sound, any sequence must have a backdrop of ambient sound if it is not to sound disconcertingly lifeless and unmoored in space. These ambient sounds tether other sounds together, allowing sound editors to bridge gaps while maintaining a sense of spatial and temporal unity. They can also be used in subtle ways to

adjust what Chion (1994:87) calls *extension*—"the degree of openness and breadth of the concrete space suggested by sound, beyond the borders of the visual field, and also within the visual field around the characters."

And so, as the tape of the conversation is played over and over again, it is accompanied by sounds of the world through which Caul moves. For the first two thirds of the film, these sounds are often ones that reinforce an association between the voices of the conversation and the equipment used to present them. Later, when the time comes to begin preparing the audience for the revelation that the voices of the conversation may be something other than a purely veridical representation of sounds on tape, the conversation is for the first time heard away from the presence of any audio device, accompanied by the sounds of Caul's footsteps as he walks across a courtyard. Finally, the ambient sounds of both Union Square and Caul's recording studio begin to escape the bounds of their representation on screen: the sound of tapes spinning audible as Caul stands in an elevator, for example. This progression lays the groundwork for the final reveal.

5.3. Visible signs of viewpoint and the final twist

A risk for all stories built on surprise and misdirection is that the audience will feel that the story itself has been (or that the people responsible for producing the story have been) a cheat. This risk is intimately tied up with issues of *responsibility* and *attribution*: who is responsible for having depicted something in a way the audience may deem inaccurate? The challenge is to find a way to avoid the appearance of inconsistency and uncooperativity, a concern that may be resolved by taking advantage of resources for representing embedded viewpoints to finesse what an ostensibly reporting speaker (or narrator) may be held responsible for with regard to those reports.

The climax of *The Conversation* is especially risky, because it involves what is manifestly a misquotation of itself—surely a violation of Grice's (1975) maxim of Quality. Furthermore, the status of the crucial snippet as represented perception is what licenses the presentation of these two different sound clips as representations of the "same" utterance. Yet it is only the fact of the surprising change in the final iteration that reveals to the viewer that the element is—or was—an instance of represented perception at all. That is impressive enough, but recall that the production history of the film (Ondaatje and Murch, 2002; Koppelman, 2005) shows that this complicated maneuver actually served to clarify previously confusing story elements for audiences.

Where earlier in the film, it was necessary to de-emphasize cues about the viewpoint status of the conversation, the opposite is the case at the moment of the climactic reveal. The audience must now be convinced beyond a doubt that the conversation represents *internal sound* (Chion, 1994) in Caul's mental soundscape: his subjective, private experience. This new unambiguous framing is not achieved through acoustic effects applied to Mark's voice to make it seem more subjective (such as slowing it down, lowering the pitch, or adding reverb). Instead it arises from the same kinds of cross-modal cues that earlier served to incorporate colorable viewpoint cues *without* drawing unwanted audience attention—laying the groundwork necessary to allow them to make sense as internal sound in retrospect—but now deployed less subtly, less ambiguously, and without distractors. Importantly, the critical swapped sequence is integrated into a larger sequence of sights and sounds that are tied firmly to Caul's viewpoint.

Caul, fearing that he has heard Mark and Ann being murdered at the Jack Tar hotel the night before, arrives at the Director's office building. He is surprised to see two things: Ann herself, sitting in the back seat of a limousine, and a newspaper with the headline "Auto Crash Kills Executive." He soon enters the reception area and sees Mark and Ann emerge together, surrounded by a mass of reporters peppering them with questions. Caul, shocked, flashes back first to the night before, the sounds of struggle now synched to images showing Mark and Ann murdering the Director at the Jack Tar, and then all the way back to the original conversation. The sequence culminates at last in Mark's "He'd kill *us* if he got the chance" (1.41′26″): new audio synched to the old image, punctuated by a final cut to the Director's corpse, underlining the connection between the two.

Two major cross-modal methods work to tie these pieces together and establish their subjective grounding in Caul's internal experience: the presence and timing of close-ups on Caul to frame each flashback, and overlapping sound across cuts. Frequent cuts to Hackman (whose acting in this scene is superb) show Caul's dawning realization and establish definitively that he is *mentally* attending to the intercut events. Meanwhile, multiple layers of overlapping sound serve to integrate tightly across three different timelines: The voices of the reporters in the present bridge the transitions to Caul's reconstructed vision of the murder, heard loudly in the foreground during the lingering close-ups on Caul's face, then faded into to the background (but not absent) following cuts to events from the night before. This ensures that the shots of the hotel retain a connection to Caul's here-and-now. Next, audio of the Union Square conversation begins to lap into the sequence from the other side: first, accompanying visuals from the hotel (beginning at 1.40′22″) and then, ahead of the cut back to Union Square, playing over the final close-up of Caul's face (1.41′09″).

Now that the apparatus of audio equipment is finally absent, the multi-layered and protracted integration of these elements makes sense far more readily as a representation of Caul's internal, subjective experience.

6. Cinematic viewpoint and the complexity of sound

How should we characterize the kinds of mixed and ambiguous viewpoint representations at play in the strategies described above, and their relationship to other kinds of viewpoint representations in language? One approach is to focus on what they do at the level of narrative structure and rhetorical impact.

In writing, as in film, it isn't always immediately clear to whom a given sequence should be attributed. Free indirect discourse and its relations lend themselves especially well to eliding or obscuring questions of attribution, and are often used in ways that capitalize on that quality. Of particular interest are the varieties that Dorrit Cohn (1966, 1978) has termed ''narrated monologue''—the presentation in third-person narrative of a character's unspoken discourse, her inner thoughts and judgments—and ''narrated perception,'' in which a character's sensory experience is presented without being explicitly marked as an act of perception. The utility of narrated monolog and represented perception for setting up surprises of the sort described here—presenting material in such a way that it seems veridical at first but later can be retroactively attributed to a limited character viewpoint—is well attested in written fiction, from detective stories (Tobin, 2009) and other genres associated with ''twists in the tail'' (Linkin, 1988) to the work of Jane Austen (Pallarés-García, 2012). In all of these cases, the trick depends on the incorporation of character viewpoint into an ostensibly narrative viewpoint, without a full, explicit shift of the deictic center, something that is possible in multiple discourse genres, including both narrative prose and film.

Charles Forceville (2002) precedes me in exploring the relationship between textual and cinematic resources for presenting ''audiovisual points of view whose origin is not a single, undividable one'' (p. 133) in light of the fact that they can be seen to serve many of the same rhetorical and narrative functions. In his study, a comparison of Ian McEwan's novel *The Comfort of Strangers* and its adaptation to film, Forceville concludes that cinematic viewpoint ''ambivalence'' and filmic methods for signaling that ''phenomena are distilled through a characters' perception'' should be analyzed ''in terms of FID'' (p. 133). This is, I think, manifestly correct—up to a point. Film makes use of its own ways of presenting mixed and ambiguous points of view to achieve many of the same wide array of discourse goals that FID serves in writing. However, we should not be too ready to set the differences in media aside, nor to consider these phenomena solely the purview of free indirect style.

In the case of *The Conversation*, the crux of the contrasting line readings of ''he'd kill us if he got the chance'' involves filmic counterparts of both direct quotation and represented perception in prose. To achieve its effects of ambiguity and misdirection, *The Conversation* relies on cinema-specific resources available for invoking—and for foregrounding and backgrounding—character viewpoint within the material affordances and editing conventions of narrative film. These include, as we have seen, (a) the continual presence of both the camera-eye and character viewpoints visible on screen; (b) the (many) circumstances under which audiences use these cinematographic viewpointings primarily as cues to orient around and over the discontinuities of cuts; and (c) the relatively underexplored contributions of sound.

In matters of viewpoint representation in multimodal discourse, sound may seem rather impoverished. Of course we know that even in speech, sound can communicate more than one thing at once, offering paralinguistic cues by way of pitch, volume, tempo, and so on. But compared to the rich and multivalent deictic structures built into the visual/ proprioceptive modality of gesture and sign, and compared to the resources they offer for representing multiple viewpoints at once through strategies like body partitioning, the communicative resources of sound for representing viewpoint can look pretty thin.

In cinema, however, the resources for representing different viewpoints at the same time are disposed differently. In the film soundtrack, any number of different sounds can be stacked on top of one another, and these overlapping sounds can be situated either at the same or at different narrative levels simultaneously—any of which can line up with or diverge with the viewpoints visible or implicit on screen. Extradiegetic orchestral music can play at the same time as dialog and sound effects synched to events visible on screen, the sounds of off-screen events, internal sounds, or voiceovers. They can even switch status mid-stream, as when music being sung or played on screen crossfades into orchestral score.

This brings us back to the enhanced coherence effect reported by Murch and Coppola. Why should it be clarifying, rather than confusing, to be presented with inconsistencies in sound?

I suggest that this outcome is consistent with the findings of Parrill et al. (2016), as well as fMRI (Zacks et al., 2010) and eye tracking (Smith and Henderson, 2008) studies indicating that continuity editing techniques (including gaze alignment and other viewpoint cues) most typically serve primarily to *smooth over* the discontinuities of edits to support event segmentation, rather than to *draw attention* to character subjectivity. If audiences tend to integrate across viewpoint shifts to understand a sequence of shots as components of a single event, backgrounded viewpoint cues will be relatively difficult to access for re-interpretation.

Under such circumstances, audiences need very salient, highly foregrounded cues drawing attention to character viewpoint to trigger the new, conscious reframing. Once the reframing *has* been triggered, however, it seems to be easily accepted. In the words of sound designer Thom (2007:124), ''. . .once the audience realizes or feels that what they are

seeing and hearing is being filtered through the brain of one or more of the characters in the film. . . then they're willing to accept almost anything you give them.''

In *The Conversation*, this is managed through the presentation of an attention-getting clash in the audio channel (the new line reading), together with both very prominent reaction shots and overlapping sound across cuts. Without the disruptive inconsistency in the line reading, Coppola found, many viewers "would just kind of tune out and say, 'Oh, they're doing that again.'" (DePalma, 1974:32). With it, audience attention is re-captured and re-directed. When the new line reading co-occurs with highly foregrounded visual cues drawing attention to Caul's viewpoint, it generates a strong new implication about the viewpoint status of what has been heard—and this new attribution can, in the manner of all frame shifts (Coulson, 2001; Coulson et al., 2006), retrospectively make new sense of latent, backgrounded elements from earlier in the discourse. In this way, the cross-modal affordances and medium-specific conventions of film allow for circumstances in which greater viewpoint complexity can, counter-intuitively, result in greater clarity and ease of comprehension.

References

Banfield, Ann, 1982. Unspeakable Sentences: Narration and Representation in the Language of Fiction. Routledge & Kegan Paul, Boston.
Barton, Stephen B., Sanford, Anthony J., 1993. A case study of anomaly detection: shallow semantic processing and cohesion establishment. Mem. Cognit. 21 (4), 477–487.
Bordwell, David, Staiger, Janet, Kristin, Thompson, 1985. The Classical Hollywood Cinema: Film Style and Mode of Production to 1960. Columbia University Press.
Chion, Michel, 1994. Audio–Vision: Sound on Screen. Ed. and Trans. Claudia Gorbman. Columbia University Press, New York.
Cohn, Dorrit, 1966. Narrated monologue: definition of a fictional style. Comp. Lit. 97–112.
Cohn, Dorrit, 1978. Transparent Minds: Narrative Modes for Presenting Consciousness in Fiction. Princeton University Press.
Coppola, Francis Ford, 1972. The Conversation, a screenplay. In: Singleton, Ralph S. (Ed.), Film Scheduling/Film Budgeting Workbook. Lone Eagle, Los Angeles,p. 1985.
Coulson, Seana, 2001. Semantic Leaps: Frame-Shifting and Conceptual Blending in Meaning Construction. Cambridge University Press, New York.
Coulson, Seana, Urbach, Thomas P., Kutas, Marta, 2006. Looking back: joke comprehension and the space structuring model. Humor 19 (3), 229–250.
Dancygier, Barbara, Lu, Wei-lun, Verhagen, Arie, 2016. In: Dancygier, Barbara, Lu, Wei-lun, Verhagen, Arie (Eds.), Viewpoint and the Fabric of Meaning. Mouton de Gruyter, Berlin.
DePalma, Brian, 1974. The Making of The Conversation: An Interview with Francis Ford Coppola. Filmmakers Newsletter, pp. 30–34.
Dudis, Paul G., 2004. Body partitioning and real-space blends. Cognit. Linguist. 15 (2), 223–238.
Emmott, Catherine, Alexander, Marc, 2010. Detective fiction, plot construction, and reader manipulation: rhetorical control and cognitive misdirection in Agatha Christie's Sparkling Cyanide. In: McIntyre, Dan, Busse, Beatrix (Eds.), Language and Style: Honour of Mick Short. Palgrave Macmillan, Houndmills, pp. 328–346.
Emmott, Catherine, Alexander, Marc, 2014. Foregrounding, burying, and plot construction. In: Stockwell, Peter, Whiteley, Sara (Eds.), The Handbook of Stylistics. Cambridge University Press, Cambridge, pp. 329–343.
Fludernik, Monika, 1993. The Fictions of Language and the Languages of Fiction: The Linguistic Representation of Speech and Consciousness. Routledge, London & New York.
Forceville, Charles, 2002. The conspiracy in The Comfort of Strangers: narration in the novel and the film. Lang. Lit. 11 (2), 119–135.
Grice, H.P., 1975. "Logic and Conversation." Reprinted in Studies in the Way of Words. Harvard University Press, Cambridge, MA 1989.
Harmon, Jim, Glut, Donald F., 1972. The Great Movie Serials: Their Sound and Fury. Doubleday, Garden City, NY.
Koppelman, Charles, 2005. Behind the Seen: How Walter Murch Edited Cold Mountain Using Apple's Final Cut Pro and What This Means for Cinema. New Riders Pearson, Berkeley, CA.
Kukkonen, Karin, 2013. Flouting figures: uncooperative narration in the fiction of Eliza Haywood. Lang. Lit. 22 (3), 205–218.
Linkin, Harriet Kramer, 1988. Narrative technique in 'An occurrence at Owl Creek Bridge'. J. Narrat. Tech. 18 (2), 137–152.
Lips, Marguerite, 1926. Le Style Indirect Libre. Payot, Paris.
McNeill, David, 1992. Hand and Mind: What Gestures Reveal About Thought. University of Chicago Press, Chicago.
Ondaatje, Michael, Murch, Walter, 2002. The Conversations: Walter Murch and the Art of Editing Film. Alfred A. Knopf, Toronto.
Pallarés-García, Elena, 2012. Narrated perception revisited: the case of Jane Austen's Emma. Lang. Lit. 21 (2), 170–188.
Pallarés-García, Elena, 2014. Narrated Perception and Point of View in the Novels of Jane Austen. Doctoral thesis.
Parrill, Fey, 2009. Dual viewpoint gestures. Gesture 9 (3), 271–289.
Parrill, Fey, Stec, Kashmiri, Quinto-Pozos, David, Rimehaug, Sebastian, 2016. Linguistic, gestural, and cinematographic viewpoint: an analysis of ASL and English narrative. Cognit. Linguist. 27 (3), 345–369.
Perniss, Pamela, Özyürek, Asli, 2008. Representations of action, motion, and location in sign space: a comparison of German (DGS) and Turkish (TID) sign language narratives. In: Signs of the Time: Selected Papers from TISLR 8. , pp. 353–378.
Pratt, Mary Louise, 1977. Toward a Speech Act Theory of Literary Discourse. Indiana University Press, Bloomington.
Rosenberg, John, 2011. The Healthy Edit: Creative Editing Techniques for Perfecting Your Movie. Focal Press, Burlington, MA.
Sanford, Anthony J., Emmott, Catherine, 2012. Mind, Brain and Narrative. Cambridge University Press, New York.
Smith, Tim J., Henderson, John M., 2008. Edit blindness: the relationship between attention and global change blindness in dynamic scenes. J. Eye Mov. Res. 2 (2), 1–17.
Sonnenschein, David, 2011. Sound spheres: a model of psychoacoustic space in cinema. N. Soundtrack 1 (1), 13–27.

Sweetser, Eve, 2012. Introduction: viewpoint and perspective in language and gesture, from the ground up. In: Viewpoint in Language: A Multimodal Perspective. Cambridge University Press, Cambridge, UK, pp. 1–24.

Teki, Sundeep, Chait, Maria, Kumar, Sukhbinder, von Kriegstein, Katharina, Griffiths, Timothy D., 2011. Brain bases for auditory stimulus-driven figure–ground segregation. J. Neurosci. 31 (1), 164–171.

Thom, Randy, 2007. Designing a movie for sound. In: Sider, Larry, Freeman, Diane, Sider, Jerry (Eds.), Soundscape: The School of Sound Lectures, 1998–2001. Wallflower Press, London, pp. 121–137.

Tobin, Vera, 2009. Cognitive bias and the poetics of surprise. Lang. Lit. 18 (2), 155–172.

Tobin, Vera, 2014. Where do cognitive biases fit into cognitive linguistics? In: Dancygier, Barbara, Borkent, Mike, Hinnell, Jennifer (Eds.), Language and the Creative Mind. CSLI Publications, Stanford, CA, pp. 347–363.

von Roncador, Manfred, 1988. Zwischen Direkter und Indirekter Rede. Mouton de Gruyter, Berlin.

Zacks, Jeffrey M., Speer, Nicole K., Swallow, Khena M., Maley, Corey J., 2010. The brain's cutting-room floor: segmentation of narrative cinema. Front. Hum. Neurosci. 4, 168.

Vera Tobin is assistant professor of Cognitive Science at Case Western Reserve University. Her previous work has included publications on viewpoint phenomena in presupposition, causal constructions, degree achievement verbs, and various kinds of irony. Her book Elements of Surprise, forthcoming from Harvard University Press, investigates the intersection of cognitive bias and the poetics of surprise. She is also the co-editor, with Fey Parrill and Mark Turner, of the volume Meaning, Form, and Body.