

# VOICE-OVER TECHNIQUE IN SPIKE JONZE'S *ADAPTATION*

SABRINA FRANCESCONI  
UNIVERSITY OF TRENTO

**Abstract** – By adopting the tools of multimodal stylistic analysis (McIntyre 2008; Nørgaard 2014; Pillière 2014; Zurru 2010), this article explores the forms and functions of the voice-over in Spike Jonze's critically acclaimed film, *Adaptation*, released in 2002. Subverting the traditional monologic and stable configuration, the film's voice-over is polyphonic and fluid: it involves multiple speakers and sometimes it features turn-taking. A polycentric semiotic technique, the voice-over inhabits the interstitial filmic spaces within the shots, as well as broader horizons beyond the shot boundaries. A range of visual-verbal patterns are also deployed, generating semiotic effects of consonance and dissonance. Results show that, rather than operating in binary opposition to the voice-in, the voice-over moves along a cline from voice-in as dialogue to the voice-over commentary typical of documentaries. As such, the aural system is also functionally layered; it serves a range of distinct yet interconnected roles, expressing thoughts, as well as passages that are read or written by the protagonist. This multifunctional resource, ultimately, unveils the collaborative and transformative process of screen adaptation narrated by the analysed meta-film.

**Keywords:** voice-over; multimodal stylistic analysis; film; adaptation; Spike Jonze.

## 1. Introduction

“And God help you if you use voice-over in your works, my friends. God help you. It's flaccid, sloppy writing. Any idiot can write voice-over narration to explain the thoughts of the character” (01:05:24-01:05:36). With these menacing words in Spike Jonze's film *Adaptation* (2002), screenwriting guru Robert McKee warns a huge audience of prospective adapters against the perils of the voice-over. In the crowded conference room sits the film's protagonist Charlie Kaufman, who is struggling to find the right ending for his script. It is only at the end of the film itself that the screenwriter declares: “I have to go right home. I know how to finish the script now. It ends with Kaufman driving home after his lunch with Amelia, thinking he knows how to finish the script. Shit, that's voice-over. McKee would not approve. How else can I show his thoughts? I don't know. Well, who cares what McKee says? It feels right. Conclusive” (01:45:06-01:45:32). Acknowledging that he is contradicting the guru's golden rule, Charlie ultimately decides to rely on the voice-over for his adaptation. Overall, the technique becomes a significant aural resource in the meta-film *Adaptation* itself (actually written by Charlie Kaufman), in terms of its extent, distribution, multimodal configurations and functions, and the unusual number of voice-over speakers.

This article seeks to explore and question the voice-over (henceforth VO) as a semiotic technique in this critically acclaimed film, overtly concerned with the challenges of adaptation. In order to unpack the VO meaning-making system, the toolkit of multimodal stylistic analysis will be adopted (McIntyre 2008; Nørgaard 2014; Pillière 2014; Zurru 2010). Attention will be devoted to the forms and functions of the VO in the film narrative and how the narrator(s), character(s) and focaliser(s) are framed through the VO (Costa 2019; Forceville 2002; Harrison 2020; Piazza 2004, 2010; Kozloff 1988). Specific attention will be given to identifying, analysing and interpreting the multimodal

dynamics inherent in the VO technique.

Results show that, in this film, the VO allows for polyphonic discourse, located within and beyond the scene, and deploys a range of multimodal configurations that generate both semiotic consonance and dissonance. In consequence, the VO is also functionally layered: it frames scenes based on thinking, remembering, writing, reading, listening. The VO ultimately unveils the complex, collaborative, and transformative nature of the adaptation process.

As for composition, this introduction is followed by a literature review. Section 3 outlines the article's methodology, Section 4 focuses on the VO technique, and Section 5 presents the film. Section 6 reports on the data collection and analysis. The last section summarises the findings and draws concluding remarks.

## 2. Literature review

Overall, the VO has been neglected in critical discussions on the meaning-making system of cinematographic discourse (Kozloff 1988, p. 2). In fact, soundtracks as a whole were initially considered theoretically irrelevant, as early film theorists including Eisenstein, Epstein, Metz, Pudovkin, Arnheim, Rotha, and Kracauer celebrated the primary visual nature of film (Kozloff 2000, p. 6). The VO was still overlooked in the 1970s, despite the fact that researchers had started to address the cinematographic soundscape (Kozloff 2000, p. 7). Studies at that time explored sound theory, technology, and quality, also in relation to “improvements in microphones, sound mixing and editing” (Kozloff 2000, p. 7). Revealing a bias in film studies against the ‘telling mode’ (Costa 2019, p. 164), verbal discourse was generally considered to be a mere accompaniment to the imagery (Kozloff 2000, p. 62; Piazza, Bednarek and Rossi 2011, p. 6).

A relatively recent topic of inquiry, filmic VO has been explored through the frame of cognitive stylistics (Harrison 2020), multimodal analysis (Piazza 2010), narratology (Costa 2019), film history and theory (Kozloff 1988, 2001). The twofold literary-linguistic approach provides evidence for the hybrid nature of the VO, often considered, celebrated or (more often) condemned as a literary device (Costa 2019, p. 167; Hutcheon 2013, p. 54; Kozloff 1988). It is not by chance that VO studies explore page-to-screen adaptations, including *The Age of Innocence* by Martin Scorsese adapting the eponymous 1920 novel by Edith Wharton (Costa 2019; Kozloff 2001), the TV series *The Handmaid's Tale*, adapting Margaret Atwood's 1985 novel with the same title (Harrison 2020) or, again, *La terra trema* (1948) by Luchino Visconti adapting *I Malavoglia* (1881) by Giovanni Verga (Piazza 2004). Arguably, the most insightful observations derive from a multimodal approach (Costa 2019; Harrison 2020; Piazza 2010), whereby the utterance is explored in its interface with other aural and/or visual semiotic systems (e.g., voice and lighting, voice and gesture). Interesting critical contributions to the study of VO have developed into a close field of inquiry, with attention to the audio-visual genre of the video documentary (Chovanec 2020; Franco 2001; Lorenzo-Dus 2009).

## 3. Multimodal stylistic analysis and filmic meaning-making

In order to unpack the meaning-making system of the VO, this paper adopts the toolkit of multimodal stylistic analysis (McIntyre 2008; Nørgaard 2010, 2014, 2019; Pillière 2014; Zurru 2010). Developed over the last two decades, this analytical framework integrates

stylistics and multimodality for text analysis and observes how different semiotic resources interact in a multimodal ensemble (Pillièrè 2014, p. 100). In his pioneering study in the field, for instance, McIntyre (2008) analyses the opening soliloquy scene from the 1995 film version of Shakespeare's *Richard III* directed by Richard Loncraine, written by Loncraine himself and by Ian McKellen, and featuring McKellen as the protagonist. The research design considers the interaction of dramatic texts and performance, with specific attention to the *mise-en-scène* elements of setting, costume, make-up, lighting, and staging. In this vein, the multimodal filmic artefact is conceived as the co-occurrence of modes and modal resources, as the visual and audio tracks are sequentially realised and interrelated, that is, unfold over time (Bateman and Schmidt 2012; Bateman, Wildfeuer and Hiippala 2017, p. 329; Toolan 2014).

The visual track encompasses dynamic images of human, natural, and cultural participants and their actions. These elements can act, move, change positions, or make gestures in the setting in which they are located, which itself can be static (e.g., a room) or dynamic (e.g., a moving vehicle). Change and motion can also be related to natural phenomena, such as a storm or a sunrise, that exhibit different durations, speeds, and rhythms. Any setting, be it natural or artificial, can serve different functions and be the site where people live, work, or practise sports. The positions participants occupy and their mutual relations within the *mise-en-scène* can reveal their interactions, as well as their degrees of relevance in an event or to the audience (Kress, van Leeuwen 2006, Chapter 4). As such, participants and actions do not have any objective form of representation but are always represented from a specific standpoint and distance. Angles (e.g., low angles, high angles, eye angles) and size of frame (e.g., close-ups, medium shots, long shots) can project the subjective dynamics of perception, which then vary as the camera moves along different trajectories (e.g., vertical, horizontal, circular), with variable durations and degrees of smoothness (Bateman, Schmidt 2012; Bateman, Wildfeuer, and Hiippala 2017, p. 329ff).

More central to the present article, the audio track includes voices, music, and sound effects. Natural or artificial, these sounds differ in their materiality: e.g., doorbells, voices, drums, car engines (van Leeuwen 1999, p. 11ff). From a socio-semiotic perspective, materiality offers a range of choices that “have semiotic value. They carry with them a potential for semiosis, for making meaning” (van Leeuwen 1999, p. 8). For example, sound designers know that the choice of strings or brass for the film soundtrack affects filmic message, tone and impact, and that these instruments interact differently with other sounds and with the visual track. Meaning is also related to the formal properties and parameters of sound—including perspective, volume, rhythm, interplay, melodies—and to the different and varying degrees of relevance these have for the public (van Leeuwen 1999). In all these components and regulated by all these parameters, the filmic soundscape is dynamic, as sounds and their relations unfold and change over time (Bordwell, Thompson 2020).

Be it speech, music or noise, filmic sound can be either diegetic or non-diegetic (Bordwell, Thompson 2020). Diegetic sound is internal to the narration, like dialogue among characters (voice-in) or noises generated by visible objects. Non-diegetic sound originates outside the story space, like the narrator's voice in a documentary video (voice-over) or the soundtrack of a film, superimposed at the post-production stage. Diegetic sounds can be on-screen or off-screen (Bordwell, Thompson 2020). If the speaker is simply excluded from the frame and s/he could be included by adjusting the camera's position, we should speak of voice-off (Kozloff 1988, p. 3). Diegetic sound can be either internal and objective or external and subjective (Bordwell, Thompson 2020). The interior monologue is an internal diegetic sound, whereby spectators are offered the thoughts and

emotions of a character. External diegetic sound is produced from a physical source within the frame like a musical instrument, a person, a vehicle. While the physical position and provenance of sound is of interest for the distinction between voice-in and VO, filmic sound can also be described in terms of its temporal relation to the depicted scene (Bordwell, Thompson 2020). Sound can be simultaneous, like voice-in within a dialogue, or non-simultaneous, like the VO of an omniscient narrator speaking from a later time than that of the story, offering a subsequent, retrospective narration.

#### 4. The voice-over

Within the filmic artefact, the VO may express the narrator(s), the focaliser(s) and/or character(s) (Harrison 2020; Piazza 2010; van Leeuwen 1991). It may encode a heterodiegetic ‘frame narrator’ speaking from a privileged vantage point outside the story boundaries (Kozloff 1988, p. 50), as in *La Terra trema* by Luchino Visconti (Piazza 2004, p. 47). Alternatively, a homodiegetic ‘embedded narrator’ may be encoded, who speaks from an internal position and expresses a subjective perspective (Harrison 2020; Piazza 2010). An example of this embedded stance is Kathy’s homodiegetic VO in *Never Let Me Go* (2010) by Mark Romanek, the adaptation of Ishiguro’s novel. A range of alternative solutions to frame and embedded narrators may be featured, depicting different degrees of social and temporal distance between the speakers and the narrated events. Indeed, the VO may belong to a character who is not the narrator. Stances and relations may also mutate across the film narrative. Moreover, there may be more than one VO speaker in a film and these different VO speakers may take on different roles and relevance within the film.

Visible or invisible, every speaker features a distinct voice quality. Film-makers strive to find the right voice quality, accent, and inflections for their themes (Kozloff 2000, p. 75). In turn, actors modulate their voices in order to express relaxed, sentimental, ironic, authoritative tones, depending on the scenes they take part in (Kozloff 1988, p. 95). Overall, the human voice can be described not only in terms of gender and age, but also according to quality, tension, roughness, breathiness, loudness, pitch range and vibrato (Barthes 1977; van Leeuwen 1999). Several of these parameters reveal the psychological and/or emotional profile of speakers: e.g., a wide pitch range expresses excitement, a narrow pitch range conveys boredom. Voice is also socio-culturally meaningful, as it reflects social or ethnic group membership (Kozloff 2000). Similarly, personal and professional experience influences both voice quality and performance, as Orson Welles’s solid radio experience affects the artist’s VO in *Citizen Crane* (Kozloff 1988, p. 33).

As for the interplay between the VO and the visual track, different patterns including overlapping, complementarity, or contradiction may generate effects like ambiguity, irony, or bewitchment (Kozloff 1988, p. 102; Kozloff 2013, p. 41). In *The Age of Innocence*, Martin Scorsese extensively relies on the VO of the actress Joanne Woodward to depict the lifestyle of the New York upper-class and to unveil the interiority of the protagonist Newland Archer (Costa 2019, p. 163). This omniscient filmic narrator, whose identity is never revealed, achieves an ironic effect “by discretely opposing what is shown by the camera” (Costa 2019, p. 174). As Costa remarks, the VO provides information or insights about characters that “unbalance spectators’ perception of facts” (Costa 2019, p. 167), without subverting them. Another example of multimodal incongruity can be detected in the first episode of the TV series *The Handmaid’s Tale*, where critically acclaimed Elizabeth Moss features the protagonist June, also known as Offred, a handmaid who is forced to bear children for wealthy families. Harrison (2020, p.

36) notes that the actress' VO, combined with visual strategies such as pervasive close-ups, colour, lighting, costumes, and symmetrical composition, has the effect of distancing June/Offred from her own tale. Through audio-visual interplay, Harrison concludes (2020, p. 25ff), the interior monologue VO creates tension rather than proximity, shaping a 'split self' that makes the speaker more a witness than an experiencer.

The VO is thus a multifunctional system (Kozloff 1988, p. 52ff), that is, it projects multiple strands of meaning. In metafunctional terms (Halliday 1978; Halliday and Hasan 1985), the VO first serves the ideational metafunction, concerned with the expression of content (human, natural, and cultural environments). If the frame narrator's VO can present characters, events, and the story's time and place, the embedded narrator tends to reveal interiority, namely thoughts, feelings, and emotions (Kozloff 1988, p. 42). The VO can also fulfil an interpersonal metafunction, focused on social relations among participants who interact in the process of film de/codification (directors, screenwriters, composers, spectators). As such, it can set a certain tone or mood and establish a connection with the audience, while sustaining the fruition and interpretation of the narrative (Kozloff 1988, p. 51). And finally, it projects the textual metafunction, involved in the construction of the text (organisation, cohesion and coherence). The textual function is related to the placement, frequency, and duration of the VO unit across the film: generally intermittent, the VO can indeed be minimally or extensively heard across the text and, as such, serve as a strategy of composition and cohesion (van Leeuwen 1991). It can then activate analeptic and proleptic leaps in time, that is, trigger flashbacks and flashforwards within the filmic narrative (Gordejuela 2019).

## 5. Case study

In light of this theoretical and methodological framework, this paper seeks to analyse Spike Jonze's film *Adaptation* (2002). *Adaptation* tells the story of a successful screenwriter, Charles Kaufman (Nicholas Cage), who tries to adapt Susan Orlean's non-fiction book *The Orchid Thief* into a film. The film's screenplay was written by Charlie Kaufman, who is also the fictional protagonist. Cage also plays Kaufman's twin brother Donald. If Charlie is introvert and has low self-esteem, Donald is extrovert and sociable. Donald, who has recently moved in with Charlie, starts working on a script and, just as Charlie and Donald's personalities differ, so do their approaches to their work and styles of writing. Charlie tries to create an original and creative artwork, without following traditional and standardised structures, whilst his twin brother relies heavily on conventional adaptation norms, following the footsteps of the screenwriting-guru Robert McKee. As Stam sustains, "together the twins manifest the split personality of many scriptwriters, torn between the art film and the blockbuster, between complexity and facile appeal" (Stam 2005, p. 2). Not only does the film highlight the role of the screenwriter in the meaning-making process enacted by the film, it also shows the various perspectives, attitudes, approaches, scriptwriters have towards the texts they want to adapt and towards the adaptation process they plan to accomplish.

Susan Orlean (Meryl Streep) is also a character in the film, her life being intertwined with that of the male protagonist of her own book: John Laroche, a rude plant-lover who tries to save rare species of orchids by stealing them. Increasingly attracted to each other, the two start a secret relationship, tainted by drug abuse, as they consume substances extracted from the rare ghost orchid. When Charlie and Donald discover this love affair and try to investigate further, they are in turn discovered by Orlean and

Laroche, who see homicide as the only possible solution to the problem. After a wild chase in a Florida swamp, Donald is first shot by the flower poacher and later hit by a car which eventually leads to his death in his brother's arms. When Laroche tries to kill Charlie as well, he is suddenly attacked by an alligator and dies. The devastated Susan Orlean is then arrested and can no longer negotiate her role as author and establish a social distance between herself and the narrative material. Ultimately, Charlie manages to finish his script and to confess his love to his long-beloved Amelia.

Robert Stam (2005) elucidates the layered conceptualisation of *Adaptation*: the film, called 'adaptation', is an adaptation and deals with an adaptation. Hence, "[t]he giddily reflexive film focuses less on the poacher than on the book's adapter struggling to write an adaptation" (Stam 2005, p. 1). In this vein, an adaptation should be considered as both product and process, as a site where formal, functional, and contextual dynamics of text codification, decodification, and recodification entwine (Hutcheon 2013). My argument is that the use of language, and of VO in particular, greatly contributes to the development of this meta-discourse. The following sections report on the data collection and analysis.

## 6. Text analysis

### 6.1. Formal patterns

There are 39 instances of VO units in the film, which last for a total of 16 minutes and 11 seconds out of 106 minutes (accounting for 15,27% of the film's total duration). VO speakers are 7 (Figure 1): Charles Kaufman (19 units, 41,5% of the VO's total duration), Susan Orlean (10 units, 34,6%), John Laroche (4 units, 16,6%), Charles Darwin (1 unit, 3,6%), Marty (1 unit, 1,5%), Robert McKee (1 unit, 0,4%), Charlie's mother (1 unit, 1,9%). Not only do the voices of the scriptwriter and the book's author resonate across the film, so do those of the agent, of the character in the book, the scientist, the scriptwriter's mother. All of these VO instances are somehow involved in the process of film adaptation. Thus, VO polyphony seems to suggest the collaborative discourse underlying the adaptation process, whose plural stances include directors, producers, distributors, music composers, set designers, screenwriters, each with their different perspectives, competences, and tools (Hutcheon 2013, p. 81).

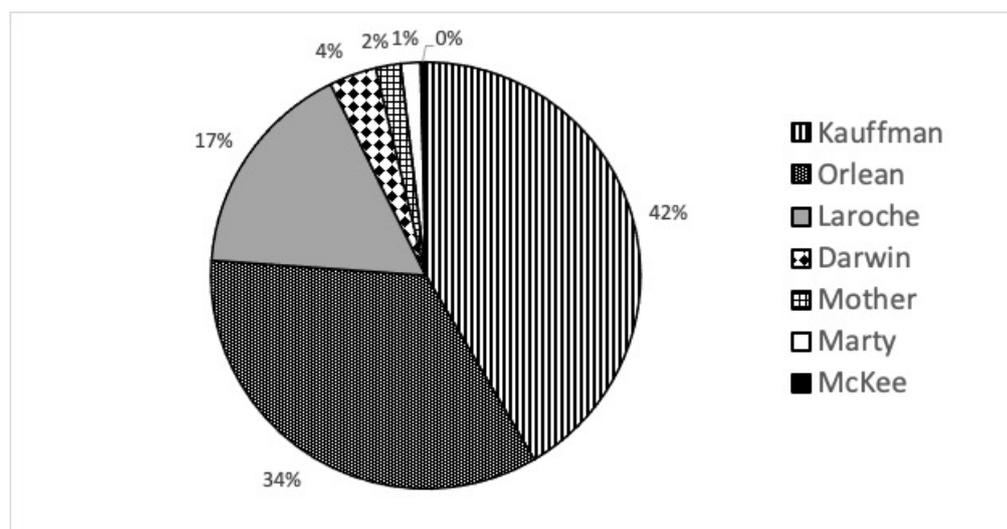


Figure 1  
VO speakers in the film and percentage of their VO speech duration.

Not only are the voices in this film adaptation plural, they also manifest as a fluid entity. In two VO units, Orlean's voice blurs into Charlie's voice; in one unit Darwin's voice blurs into that of Kaufman. For example, (1) is set at the orchid exhibition. Through a long shot, we first see Charlie walking across the exhibition area and admiring an impressive range of flowers all around him. We also enjoy numerous subjective close-up shots of the particular orchid he is observing. Susan's VO offers several similes comparing orchids to animals, vegetables, human beings. After 15 seconds (marked with > in the quote), Susan's voice mutates into Charlie's voice:

(1) There are more than 30,000 known orchid species. One looks like a turtle. One looks like a monkey. One looks like an onion. > One looks like a schoolteacher. One looks like a gymnast. One looks like that girl in high school with creamy skin. One looks like a New York intellectual with whom you do the *Sunday Times* crossword puzzle in bed. One looks like a Midwestern beauty queen. One looks like Amelia. One has eyes that dance. One has eyes that contain the sadness of the world (00:34:44-00:35:27).

In this passage, orchids are personified and attributed physical, emotional, and biographical traits. The VO is used to describe the profound connection between the human and vegetable worlds. Echoing the visual effect of a cross-dissolve, this progressive aural transition seems to suggest the absence of neat boundaries between authorial voices in the adaptation process, enacted by the ceaseless dialogue between texts, authors, and characters. Even if passages from the adapted work are reported verbatim in the adaptation, choices in terms of selection, placement, interplay with images and music entail a slow but clear transformation (Hutcheon 2013).

As for composition, the VO operates at the intra-shot and inter-shot levels, within and across shot boundaries. The aural technique generally corresponds to a single and distinct scene, featuring a recognisable space-time unit. This is the case with the predominant interior monologue VO unit, like (2). Depressed and confused, Charlie is standing alone, against the wall outside the studio where the film is being shot:

(2) What am I doing here? Why did I bother to come here today? Nobody even seems to know my name. I've been on this planet for 40 years, and I don't understand a single thing. Why am I here? How did I get here? (00:02:10-00:02:27)

The scene depicts Charlie's sense of alienation in relation to the film industry, which seems to ignore his presence and role. The internal intra-shot VO gives expression to Charlie's thoughts in the particular situation in which he is immersed.

Sometimes, however, the VO overcomes the borders of the scene and corresponds to a more complex multiple-scene unit, actually a filmic sequence. For example, the VO in (3) starts in Susan's office, while she is writing her volume. Her VO utters: "Orchid hunting is a mortal occupation", then her words trigger the text she is writing, announced through the first of three intertitles: Orinoco River, Venezuela, One Hundred Years Earlier (00:14:59)

(3) Victorian-era orchid hunter William Arnold drowned on a collecting expedition. Ormers vanished without a trace in Asia. Augustus Margary survived toothache, rheumatism, pleurism and dysentery only to be murdered when he completed his mission and travelled beyond Bhamo. Laroche loved orchids, but I come to believe he loved the difficulty and fatality of getting them almost as much as he loved the orchids themselves (00:14:53–00:15:53).

All the different cases (Arnold, Ormers, Margary) mentioned in Susan's voice and depicted in black and white are localised (Venezuela, Borneo, China) through the intertitles. If visual intertitles signal the contextual specificity of the three stories, the aural VO marks their common narrative thread. Operating as a frame narrator, the voice acts as a cohesive device across the complex sequence, holding the different fragments/stories together and organising the narrative.

If the VO is clearly an aural semiotic system, a multimodal approach is necessary for its identification, analysis and interpretation. This statement is in line with claims by Harrison (2020) and Piazza (2004). In the film's actual system, the VO acts through a range of diverse multimodal configurations, in relation to the integration of the visual and the audio tracks. Instances seem to generate consonant or dissonant interplay (Bateman, Wildfeuer, and Hiippala 2017). Consonance is expressed when the VO belongs to the person visible on screen, that is, when we hear the voice of a person who is either thinking or writing. Dissonance is expressed when the speaker is not visible, either because a different person is on screen or because an action is being displayed. These modes of interplay will be discussed in the next section, since they have a functional role in the filmic artefact. Hence, after considering the frequency and duration of VO units; the number, status, and interplay of speakers; the intra-shot or inter-shot extent of the VO segment, the following section will address the functional value of the technique within the film *Adaptation*.

## 6.2. Functional patterns

In the film, the formally diverse VO units serve different functions (Figure 2), in relation to the activities the speaker is accomplishing: thinking (44%), writing (20%), reading (8%), listening (18%).

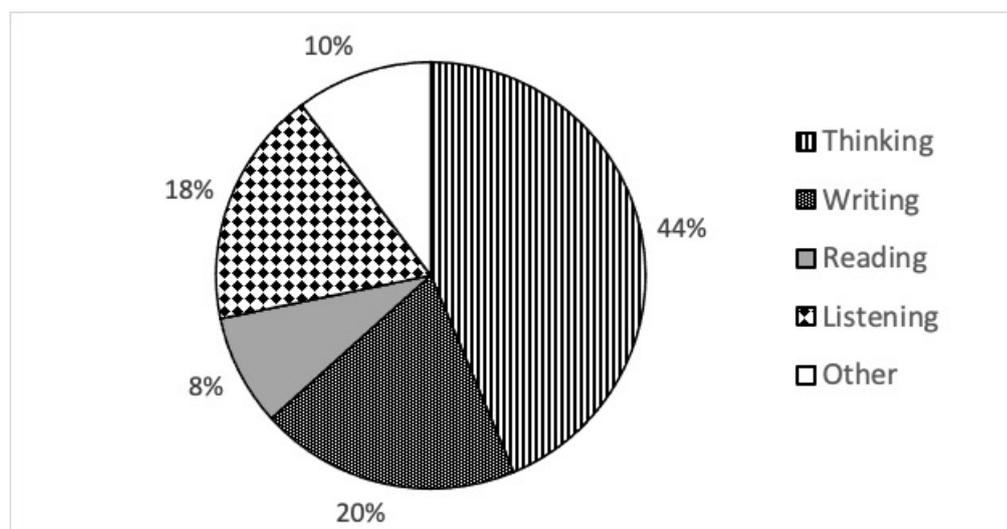


Figure 2  
VO functions in the film.

Predominantly, the internal diegetic VO expresses the inner thoughts and feelings of the protagonist and operates at the intra-shot level. For example, in (4) Charlie Kaufman is sitting at his desk, in a poorly-lit room, trying to start his script. He is first shown through a long shot from behind, then through a medium shot from the side. Then we cut to the white sheet in the writing machine as a PoV shot: Charlie's gaze becomes our lens. Then, moving back to the side perspective angle, the *mise-en-scène* frames both Charlie and the writing machine, on which he puts his fingers, without writing anything. His interior monologue VO expresses his thoughts, feelings, sensations in a fractured and disconnected way, his plan being continuously interrupted and disturbed by other ideas:

(4) To begin, to begin, how to start. I'm hungry. I should get coffee. Coffee would help me think. But I should write something first, then reward myself with coffee. Coffee and a muffin. Okay, so I need to establish the themes. Maybe banana-nut. That's a good muffin (00:14:26-00:15:51).

From a multimodal perspective, the speaker is indeed on screen, yet he is not speaking (in this case there would be a voice-in) but thinking. This unit epitomises the frequent interior monologue VO technique, with long passages of uninterrupted direct thought or feelings (Alber 2017, p. 280). In line with Harrison (2020, p. 36), the internal diegetic VO in this film encodes a kind of split self-representation, highlighting Charlie's ambivalent approach to screenwriting.

An extreme and exceptional instance of the thinking VO opens the film (00:00:34-00:01:00). As Stam writes, *Adaptation* "begins with Charlie's voice-over ruminations superimposed on a dark screen" (Stam 2005, p. 2). Since the whole unit has 326 words, the first section is sufficient to give an idea of how it functions:

(5) Do I have an original thought in my head? My bald head? Maybe if I were happier, my hair wouldn't be falling out. Life is short. I need to make to most of it. Today is the first day of the rest of my life. I'm a walking cliché. I really need to go to the doctor and have my leg checked. There's something wrong. A bump. The dentist called again. I'm way overdue. If I stopped putting things off, I'd be happier. All I do is sit on my fat ass. If my fat ass wasn't fat, I'd be happier (00:00:34-00:00:40).

Viewers are not only informed of Charlie's thoughts, but experience immersion in his entangled mind. Against that black backdrop and through multimodal dissonance, the

voice quality and the VO message are enhanced. Charlie's voice seems to originate from an abyss rather than from Charlie's body. Dissonance is unusual in the interior-monologue VO, while it is predominant when the VO is used to trigger and express a flashback. Both situations may fall under the 'thinking' category, but their temporal relation to the depicted scene is different. If the interior monologue entails simultaneity, the flashback implies retrospection (Alber 2017; Gordejuela 2019). From the viewpoint of multimodality, the interior-monologue VO shows the thinker, while the flashback depicts the object of remembrance, the memory itself. The systemic adoption of the interior-monologue VO rather than the flashback seems to frame the meta-narrative dimension of the film, which is less concerned with retelling the story than with revealing the emotional and cognitive forms of engagement with the story being retold.

The VO often expresses the process of text writing, as in (6). First, we see the building that houses the *New Yorker* magazine at night, three years later, as the intertitles state. After an extremely long shot, we fade to a closer shot showing some flats, and we zoom in to Orlean's office, where she is writing. We cut to some books on orchids surrounding her. The camera slowly moves in to show the books, some notes, a picture of Laroche, then Susan herself, active, focused and productive in front of her computer. The whole unit is accompanied by Susan's VO:

(6) John Laroche is a tall guy, skinny as a stick, pale-eyed, slouch-shouldered, sharply handsome, despite the fact he's missing all his front teeth. I went to Florida two years ago to write a piece for the *New Yorker*. It was after reading a small article about a white man and three Seminole men arrested with rare orchids they'd stolen out of a place called the Fakahatchee Strand State Preserve (00:06:12-00:06:35).

As above, the speaker is visible on screen, but is writing and not speaking her text. An example of the VO in a writing scene configuring multimodal dissonance can be found in (3).

In turn, Orlean's text is read by Kaufman in unit (7). Taken from a frontal close-up, Charlie is in his depressing dining room, sitting at the table: he has a plate with some salad in front of him, but his attention is captivated by the book he is reading and holding in his right hand.

(7) Orchids are the sexiest flowers on earth. The name orchid derives from the Latin orchis, which means testicle (00:21:43-00:21:50).

In this visual-verbal tension, Susan's VO gives aural expression to her text, which makes meaning when visually read by Kaufman. From the visual perspective, we see Charlie reading Susan's book; from an aural perspective, we hear Susan reading her book. The interconnection between codification and decodification in the process of adaptation is highlighted by Linda Hutcheon (2013, p. xv) who observes that the term 'adaptation' is used "to refer to both a product and a process of creation and reception". This twofold meaning is depicted in (1) when Orlean's VO mutates into Charlie's. If the aural shift in (1) is sequential, with Charlie's voice following that of Susan, involvement in (7) is simultaneous, instantiated in the reading process through multimodal dissonance.

A rarer function of the VO is related to processes of listening. VO unit (8) offers an unusual multimodal solution, whereby Charlie listens to his own recorded voice:

(8) [...] the insects, the mammals, the primates, the monkeys. The simple monkeys. Old-fashioned monkeys giving way to the new ones. Whatever. And then apes. Whatever. And man. Then we see the whole history of human civilization: hunting, love, war, heartache, disease, loneliness, technology. And we end with Susan Orlean in her office at the *New Yorker*

writing about flowers, and bang! The movie begins. This is the breakthrough I've been hoping for. It's never been done (00:40:44–00:41:09).

After his long writer's block, Kaufman had suddenly got some good ideas for the script and captured them on his tape recorder. In this scene, he is lying on his bed and listening to his own recorded voice. The narrative strategy inscribes the tension between the (non)simultaneity of the vocal utterance and its decodification.

A more frequent cinematographic solution is the VO typology related to phone calls (Charlie calls his mother, Susan calls John, Marty calls Charlie), especially through reverse shots, and message recordings. For example, in (9), Charlie is listening to a phone message being left by his agent Marty:

(9) Hey, superstar. It's Marty, super agent. I just wanna remind you it's been 13 weeks and Valerie's anxious to see a draft. So if you could wrap things up and get it to her by Monday, that's be great. Call me when you get this. Adios amigo (00:45:04–00:45:18).

Captured by a frontal shot, his legs crossed, the screenwriter is sitting on the messy floor of his bedroom, surrounded by books, maps, pens, pencils, cups. As in (4), the *mise-en-scène* frames the numerous sources the screenwriter relies on in the writing process. While listening to the message, Charlie closes the book he is reading, a concerned expression on his face. As above, the text codifier can be heard but not seen, while the text decodifier can be seen but not heard, through multimodal dissonance. This kind of speaker and this kind of utterance may add a logistic and pragmatic component to the story of an adaptation. Inspiration, aesthetics, and narratology need to come to terms with deadlines and pressure, as well as with the constantly "panicked and sweating" states (Stam 2005, p. 1) experienced by the adapter.

The following Table sums up the multimodal configuration of VO scenes and their meaning-making within the filmic narrative.

Narrative function	Verbal track	Visual track	Multimodality
Thinking	Speaker's VO	Speaker	Consonance
Thinking	Speaker's VO	Black screen or object	Dissonance
Writing	Writer's VO	Writer	Consonance
Writing	Writer's VO	Narrative	Dissonance
Reading	Reader's VO	Reader	Consonance
Reading	Writer's VO	Reader	Dissonance
Listening	Speaker's VO	Listener (as speaker)	Consonance
Listening	Speaker's VO	Listener	Dissonance

Table 1  
The VO Multimodal Configuration in the Film.

Overall, the main functions of the VO technique identified in this analysis cast light on the different stages of the screenwriting process: listening, reading, thinking, storytelling, and writing. Subject to decodification and codification, the story to be adapted is remembered, told and retold across time, space, situations, genres, and modes. Financial, legal, and marketing dynamics proper to the adaptation industry, which regulate the page-to-screen transformation, are also at stake. All of these activities have emotional and cognitive dimensions, and demand different degrees and different forms of involvement (Hutcheon 2013). As the alternation between multimodal consonance and dissonance showcases, the concepts of author and text are questioned in the adaptation process. Ultimately, scriptwriting is challenged and deconstructed in its layered, contextual, multifaceted aspects.

## 7. Conclusions

By adopting the tools of multimodal stylistic analysis, this article has explored the forms and functions of the VO in the film *Adaptation* by Spike Jonze. Results show that VO is used in 15,27% of the film and that seven diverse VO speakers are featured across the narrative, but the most prevalent one is the protagonist (41,6% of the film's total VO speech). Charles Kaufman plays the screenwriter, but his status and role are far from neat and stable: he is mainly represented while thinking (accounting for 44% of the duration of all VO scenes), as well as while remembering, writing, reading, listening. In all these scenes, the VO is systemically used to challenge a simplistic representation of screenwriting.

A range of multimodal configurations are enacted in the VO units, in both simple scenes and in more complex sequences, generating semiotic visual-verbal consonance and dissonance. Rather than featuring VO as the clearly defined strategy described by film manuals – neatly distinct from voice-in, a matter of binary choice – these VO units seem to trace a continuum, with the VO commentary by an invisible storyteller at one extreme end of a cline and the voice-in as dialogue at the other end of the cline. Hence, these VO instances seem to operate according to a graded system, with matters of degree being determined by multimodal patterns, temporal relations, narration stances and functions. In the film, the multimodal and parametric configuration of the VO shows that the adaptation process requires and entails collaboration, integration, tension and transformation.

This film may also be considered in relation to the controversial position the VO holds within adaptation studies (Hutcheon 2013, p. 54). Arguably, literary scholars who are sensitive to fidelity discourses celebrate the audio resource as a pure technique that enables the film adaptation to 'faithfully' transpose the adapted literary text (e.g. through word-by-word transfers of interior monologues or of narrative segments that provide temporal contextualisation or summaries). For the same reason, the VO is often criticised by film scholars for being a linear transfer of a 'literary' form of verbal expression, thus anchoring the filmic artefact in the semiotic resources of the literary work, and not in the affordances of the audio-visual medium. Subverting such aprioristic and essentialist stances, as well as simplistic definitions, this film shows that the VO is actually multifaceted and multifunctional.

The predominant interior-monologue function of the VO seems to foreground the meta-filmic dimension of *Adaptation*. Here, the VO is less about the story being told than the process of story-telling, as revealed through its psychological, socio-cultural, and economic facets. Central to this meta-filmic dimension, the VO is used to question and challenge the process of adaptation from a fidelity lens. It does not transfer passages of the adapted text, but expresses the dynamics, difficulties, and limitations of the adaptation process. Meanwhile, and as a reaction to the position shared by many film scholars, which claims the filmic medium is primarily visual, this text showcases the semiotic relevance of the audio track, and specifically the audio device of the VO for meaning-making. However, since the VO can only be identified and comprehended from a multimodal perspective, *Adaptation* highlights the irreducibly multimodal dimension of the filmic artefact. The film cannot be defined and perceived as primarily visual but is better understood as an audio-visual text that makes meaning through the integration of semiotic resources.

This paper has various limitations related to the case study and the research questions being raised. First, it would be more illuminating to closely explore the VO in relation to other filmic sound systems such as voice-in and music, as well as to visual

modes. Second, examining a large number of audio-visual texts would provide a clearer picture of the forms and functions of the VO in a film's meaning-making system and test the validity of the hypotheses formulated here. Third, a contrastive investigation should be carried out on film adaptations and video documentaries. Future research should thus explore the VO technique in more detail as well as more broadly, in order to offer a fuller picture of the role and influence of the VO in audio-visual texts.

**Bionote:** Sabrina Francesconi is Associate Professor of English Linguistics at the University of Trento. Her research interests are tourism and heritage discourses, inspected through multimodal genre analysis, and Alice Munro's style, explored through systemic-functional stylistics. Recent publications include the articles "Film Adaptations as Intersemiotic Contact Zones: *Edge of Madness* by Anne Wheeler" (*Textus*, 2019) and "Short-film Adaptations of Munro's Stories in the 1980s: A Multimodal Stylistic Analysis" (*Lingue e linguaggi*, 2019), as well as the monographic volumes: *Heritage Discourse in Digital Travel Video Diaries* (Tangram, 2018), *Alice Munro, il piacere di raccontare* (Carocci, 2015) and *Reading Tourism Texts: A Multimodal Analysis* (Channel View Publications, 2014). She is currently working on screen adaptations of Munro's short stories.

**Author's address:** [sabrina.francesconi@unitn.it](mailto:sabrina.francesconi@unitn.it)

## References

- Alber J. 2017, *The Representation of Character Interiority in Film: Cinematic Versions of Psychonarration, Free Indirect Discourse and Direct Thought*, in Hansen P.K., Pier J., Roussin P. and Schmid W. (eds.), *Emerging Vectors in Narratology*, de Gruyter, Berlin, pp. 265-283.
- Barthes R. [1961] 1977, *The Grain of the Voice*, in *Image, Music, Text*, Hill and Wang, New York, pp. 179-189.
- Bateman J. and Schmidt K.H. 2012, *Multimodal Film Analysis: How Films Mean*, Routledge, London/New York.
- Bateman J., Wildfeuer J. and Hiippala T. 2017, *Multimodality. Foundations, Research and Analysis. A Problem-Oriented Introduction*, De Gruyter, Berlin.
- Bordwell D. and Thompson K. 2020, *Film Art: An Introduction*, McGraw-Hill, New York.
- Chovanec J. 2020, *Voice-over and Presenter Narration in TV Documentaries*, in Hofmann C. and Kirner-Ludwig (eds.), *Telecinematic Stylistics*, Bloomsbury, London, pp. 141-164.
- Costa C.B. 2019, *An Ironic Overlay: The Use of Voice-over Narration in The Age of Innocence*, by Martin Scorsese, in "Adaptation" 13 [2], pp. 161-175.
- Forceville C. 2002, *The Conspiracy of The Comfort of Strangers: Narration in the Novel and in the Film*, in "Language and Literature" 11 [2], pp. 119-135.
- Franco E.P.C. 2001, *Voiced-over Television Documentaries: Terminological and Conceptual Issues for the Research*, in "Target: International Journal of Translation Studies" 13 [2], pp. 289-304.
- Gordejuela A. 2019, *Understanding Retrospection: Blended Joint Attention and Multimodal Construction of Film Flashbacks*, in "Language and Literature" 28 [2], pp. 115-132.
- Halliday M.A.K. 1978, *Language as Social Semiotics: The Social Interpretation of Language and Meaning*, Edward Arnold, London.
- Halliday M.A.K. and Hasan R. 1985, *Language, Context and Text. Aspects of Language in a Social-semiotic Perspective*, Oxford UP, Oxford.
- Harrison C. 2020, *'The truth is we're watching each other': Voiceover Narration as 'Split Self' Presentation in The Handmaid's Tale TV Series*, in "Language and Literature" 29 [1], pp. 22-38.
- Hutcheon L. and O'Flynn S. 2013, *A Theory of Adaptation*, Routledge, London/New York.
- Jonze S. 2002, *Adaptation* (film), Columbia Pictures.
- Kozloff S. 1988, *Invisible Storytellers: Voice-Over Narration in American Fiction Film*, University of California Press, Berkeley.
- Kozloff S. 2000, *Overhearing Film Dialogue*, University of California Press, Berkeley.
- Kozloff S. 2001, *Complicity in the Age of Innocence*, in "Style 2, Women as Narrators", 35 [2], pp. 270-288.
- Kozloff S. 2013, *Further Remarks on Showing and Telling*, in "Cinema Comparative Cinema" 1 [3], pp. 36-45.
- Kress G. and van Leeuwen T. 2006, *Reading Images: The Grammar of Visual Design*, Routledge, London/New York.
- Lorenzo-Dus N. 2009, *Television Discourse: Analysing Language in the Media*, Macmillan, Basingstoke.
- McIntyre D. 2008, *Integrating Multimodal Analysis and the Stylistics of Drama: A Multimodal Perspective on Ian McKellen's Richard III*, in "Language and Literature" 17 [4], pp. 309-334.
- Morano R. 2017, *The Handmaid's Tale* (TV Series 1), Episode 1.
- Nørgaard N. 2010, *Multimodal Stylistics: The Happy Marriage of Stylistics and Semiotics*, in Hamel S.C. (ed.), *Semiotics Theory and Applications*, Nova Science Publishers, New York, pp. 255-260.
- Nørgaard N. 2014, *Multimodality and Stylistics*, in Burke M. (ed.), *The Routledge Handbook to Stylistics*, Routledge, London/New York, pp. 471-484.
- Nørgaard N. 2019, *Multimodal Stylistics of the Novel, More Than Words*, Routledge, Abingdon, Oxon.
- Piazza R. 2004, *Features of the Authorial Voice-over in La terra trema*, in "The Italianist" 24 [1], pp. 47-62.
- Piazza R. 2010, *Voice-over and Self-narrative in Film: A Multimodal Analysis of Antonioni's When Love Fails (Tentato Suicidio)*, in "Language and Literature" 19 [2], pp. 173-195.
- Piazza R., Bednarek M. and Rossi F. 2011, *Telecinematic Discourse, Approaches to the Language of Films and Television Series*, John Benjamins, Amsterdam/Philadelphia.
- Pillièrre L. 2014, *Crossing New Frontiers? Investigating Style from a Multimodal Perspective*, in "Études de stylistique anglaise" 7, pp. 99-120.
- Romanek M. 2010, *Never Let Me Go* (film), DNA Films & Film4 Productions.
- Scorsese M. 1993, *The Age of Innocence* (film), Columbia Pictures.
- Simpson P. 2000, *Language, Ideology and Point of View*, Routledge, London/New York.
- Stam R. 2005, *Introduction: The Theory and Practice of Adaptation*, in Stam R. and Raengo A. (eds.)

- Literature and Film: A Guide to the Theory and Practice of Film Adaptation*, Blackwell, Oxford, pp. 1-52.
- Toolan M. 2014, *Stylistics and Film*, in Burke M. (ed.), *The Routledge Handbook to Stylistics*, Routledge, London/New York, pp. 455-470.
- van Leeuwen T. 1991, *Conjunctive Structure in Documentary Film and Television*, in "Media/Discourse" 5 [1], pp. 76-114.
- van Leeuwen T. 1999, *Speech, Music, Sound*, Macmillan, London.
- Visconti L. 1948, *La terra trema* (film), Universalialia.
- Zurru E. 2010, *Breaking the Crystal Shoe: A Multimodal Stylistic Analysis of the Character of Cinderella in the Shrek Saga*, in "Textus: English Studies in Italy" 23 [1], pp. 235-261.