

Multimodal stance and engagement in digital video methods articles

Christoph A. Hafner

City University of Hong Kong (China)

elhafner@cityu.edu.hk

Abstract

The proliferation of digital media technologies has led to fundamental changes in the way that we communicate, changes that have also been felt in the realm of scholarly communication. One underresearched scholarly digital genre is the “video methods article” (VMA) in experimental science, which is published by the *Journal of Visualized Experiments (JoVE)*, and whose purpose is to share advances in scientific methods with members of the scientific community. The genre draws on the medium of digital video in order to communicate new methods through multiple modes (e.g., spoken, written and visual), making it possible for scientists not only to read about but also to see new scientific methods as they are demonstrated on screen. In addition, the genre opens up possibilities for interpersonal engagement with the audience (such as the ability to speak directly to the camera) that are not present in traditional methods articles. This article draws on a corpus of 11 VMAs (1 per year from 2006 to 2016) in order to provide a multimodal analysis of key sections. It aims to show how stance and engagement are realized in VMAs through a complex multimodal interplay constructed by multiple individuals. Semiotic resources identified include elements of the researcher’s video recorded performance such as speech, gesture, facial expression, gaze, dress, and body; elements of the setting, such as chosen location, represented human and non-human participants and represented action; use of scientific visuals and animations; filmic elements such as camera angle, movement, and distance.

Keywords: English for specific purposes, languages for specific purposes, scholarly communication, digital media, digital video, multimodality

Resumen

La expresión multimodal de la posición y la implicación en el vídeo-artículo de métodos

La proliferación de las tecnologías digitales ha provocado cambios fundamentales en nuestra forma de comunicarnos, cambios que también se han dejado sentir en el ámbito de la comunicación académica. Un género académico digital poco estudiado es el vídeo-artículo de métodos (*video methods article*, VMA) en ciencias experimentales, publicado por el *Journal of Visualized Experiments (JoVE)*, cuyo objetivo es compartir los avances en métodos científicos con los miembros de la comunidad científica. Este género se basa en el medio del vídeo digital para comunicar nuevos métodos a través de múltiples modos (ej. hablado, escrito y visual), haciendo posible que los científicos no solo lean sobre dichos métodos, sino que también los vean tal y como se muestran en la pantalla. Además, el género abre posibilidades de implicación interpersonal (*engagement*) con el público (como la posibilidad de hablar directamente a la cámara) que no están presentes en los artículos sobre métodos tradicionales. Este artículo se basa en un corpus de 11 VMAs (uno por año entre 2006 y 2016) para ofrecer un análisis multimodal de secciones clave. Su objetivo es mostrar cómo la posición (*stance*) y la implicación se materializan en los VMAs a través de una compleja interacción multimodal construida por múltiples individuos. Entre los recursos semióticos identificados se incluyen elementos de la actuación grabada del investigador, tales como el habla, los gestos, la expresión facial, la mirada, la vestimenta y el cuerpo; elementos del escenario, como el lugar elegido, los participantes humanos y no humanos representados y la acción representada; el uso de elementos visuales y animaciones científicas; o elementos fílmicos como el ángulo, el movimiento y la distancia de la cámara.

Palabras clave: inglés para fines específicos, lenguas para fines específicos, comunicación académica, medios digitales, multimodalidad

1. Introduction

This article engages with two main areas of concern for ESP researchers and practitioners. The first relates to an ongoing interest in the innovation and evolution of forms of representation in digital media, as new genres are adopted in specialized communication. Digital media make available affordances for meaning making, such as hypertext, multimedia, and digital interactivity. In addition, they also make it possible for specialists to connect with more diverse audiences including both specialists and non-specialists, for example in social media or through blogs. There is now a considerable amount of research that examines the effect of such innovations and the emergence of digital, multimodal genres in academic and professional settings, with a number of special issues and volumes dedicated to this topic (e.g., Kuteeva & Mauranen, 2018; Luzón & Pérez-Llantada, 2019, 2022;

Hafner & Pun, 2020; Hafner et al., 2023). This research shows how the contemporary communicative landscape requires specialists to navigate an expanded repertoire of genres, as they go about their professional lives.

The second area relates to a growing interest in multimodal forms of expression in specialized discourse. This includes a focus on multimodal “artifacts” like webpages, blogs, social media posts, graphical abstracts, and so on. It also includes a focus on multimodal “performances” like lectures, conference presentations and their ensuing Q&A sessions. Such research is motivated by the observation that communication is always achieved through multiple semiotic resources. Thus, a focus on language alone is insufficient to explain how meaning is made in specialized genres and multimodal (interaction) analysis is required. Digital video genres are a particularly interesting focus of study because of the way in which they marry the two areas of interest outlined above. They can be analyzed as both “artifact” and “performance”.

In this article, I examine an under-researched digital video genre in the scholarly domain, namely, the video methods article (VMA) (Hafner, 2018). The analysis aims to better understand the range of semiotic resources involved in the expression of stance and engagement (Hyland, 2005). It also seeks to expand on existing research by responding to Xia’s (2023a) call to engage more with embodied forms of expression like gesture and facial expression in multimodal analysis of digital video genres. The article begins by reviewing relevant scholarship on language and digital media in scholarly communication. It then examines the concepts of stance and engagement and how these can be multimodally co-constructed. Finally, it applies these concepts in the analysis of a small corpus of VMAs.

2. Language and digital media in scholarly communication

Pérez-Llantada (2016) identifies digital media in scholarly communication as one area in need of attention in research on languages for specific purposes (LSP). In this context, digital media provide a number of affordances that can transform the nature of scholarly communication, fostering genre innovation. As already noted, these affordances include innovative ways of making meaning, as digital media facilitate multimodal, hypertextual forms of expression. In scholarly contexts, this can lead to more widespread

adoption of visual genres like the graphical abstract (Sancho Guinda, 2015). Digital media also allow for easy distribution of “content” through the Internet, enabling interaction with diversified audiences including both specialists and non-specialists, as in academic blogs (Luzón, 2013) and social media posts (Luzón & Alberro-Posac, 2020; Luzón, 2023; Tardy, 2023). These affordances facilitate collaborative relationships between specialists and non-specialists (McGrath, 2016) and also make it possible for scholars to engage with the general public in activities like citizen science (Reid & Anson, 2019) and popularization (Xia, 2023a). Thus, literacy practices in digital media are multimodal and polycontextual, frequently demanding attention to multiple audiences simultaneously and blending written and spoken styles in a way that necessitates the adoption of hybrid genres.

The developments in digital media noted above have been accompanied by shifts in social values and practices, meaning that emerging, digital genres are shaped by both social exigences and technological affordances and constraints. For example, Rowley-Jolivet and Carter-Thomas (2019) point to a trend to reduce scientific work into “scholarly soundbites” that coincides with the “media logics of marketization, popularization, immediacy, and the pursuit of publicity and visibility” (p. 83). Such values tend to be present in emerging presentation genres like the three-minute thesis (3MT) and TED Talks, which also respond to an increased competitiveness in academic and higher education contexts (Hyland & Zou, 2022). Similarly, the Open Science movement embodies values of greater transparency in academic research, with related innovations such as academic blogs (Luzón, 2013), open science notebooks (Carter-Thomas & Rowley-Jolivet, 2017) and open reviews (Breeze, 2019). Such innovations make use of digital tools for processes of creation, online dissemination, and online interaction, bringing an unknown online audience into the picture, adding to the rhetorical complexity and hybridity of online genres.

Recent work that examines emerging genres from a linguistic point of view finds that the use of language differs from that in more traditional academic genres, like research articles. Composing for a wider, more diverse audience brings with it particular rhetorical challenges, such as how to engage and persuade this less predictable audience, while making use of genres that are often associated with everyday (rather than academic) communication. Taking academic blogs as a case in point, Zou and Hyland (2020) show that academic bloggers perceive the need to engage readers by being “more personal, attitudinal, sparkling and direct in blogs” (p. 290). This is realized

by greater reliance on linguistic features, such as explicit mention of self (“I”) and others (“you”), use of attitude markers to evaluate ideas, use of boosters to be assertive, and use of questions to encourage interaction, for example. At the same time, there remains a scholarly concern to state claims cautiously through the use of hedging devices, a feature which underscores the hybridity of this genre.

As well as “written” genres like the academic blog, there has also been considerable interest in what we can term performance genres, like the 3MT and TED talks. Both of these genres consist of video recorded presentations (a kind of rehearsed performance) that are post-processed by a production team before being shared with a diverse, online audience. A number of studies have examined the linguistic features of these genres (e.g., Scotto di Carlo, 2014; Hyland & Zou, 2022). However, both genres also provide a rich array of multimodal semiotic resources that can be exploited by presenters, including visuals on powerpoint slides and the use of the body with speech, gesture, facial expression, proxemics, and gaze. In the case of TED talks, a number of recent studies (Harrison, 2021; Xia & Hafner, 2021; Bernad-Mechó & Valeiras-Jurado, 2023; Xia, 2023a, 2023b) have therefore adopted multimodal analysis, in order to provide a richer picture of how meaning is made. The present study adopts this multimodal approach in examining a genre designed for a specialist audience (the video methods article), focusing especially on interpersonal strategies of stance and engagement in the genre.

3. The video methods article genre

Hafner (2018) provides an exploratory genre analysis of video methods articles (VMAs). VMAs are a kind of scientific, video genre, whose purpose is to demonstrate (by verbally describing and by visually showing) innovative methods to the scientific community. The genre was pioneered by the *Journal of Visualized Experiments (JoVE)*, founded in 2006, in response to the replication crisis, the problem that it can be difficult to reproduce some experimental procedures based on the written protocols that appear in traditional research articles. Interviews with members of the scientific community indicate that these videos are designed with a specialist audience in mind, unlike some other video genres like TED Talks and video abstracts. While VMAs started out as amateur videos filmed by scientists in their own

lab, they very quickly evolved to become a professionally produced text with high production values: *JoVE* now maintains production teams in a number of cities around the world, who will assist scientists in the production of their VMAs, getting involved in script writing, filming, and editing.

As a professionally produced genre that draws on the skills of a production team as well as the expertise of the contributing scientists, the VMA genre is a *co-constructed* one. This adds an interesting dimension, one that is also present in other kinds of scientific video genres like TED talk videos (Xia & Hafner, 2021; Xia, 2023b): namely, authorship is highly distributed. Scientists initially contribute a written text for review, in the form of a written methods article. Once this has passed review, the production team work with the scientists to create a script, film the scientists in the lab, and edit the video, drawing on and adapting visuals provided by scientists. Hafner (2018) suggests that Goffman's (1981) notion of "production format" can help to understand this process of co-construction. According to this framework, the "production of text can be divided between a principal, whose beliefs and views are represented by the words uttered; an author, who composes those words; and an animator, who acts as the 'sounding box' to make the utterances" (Hafner, 2018, p. 25). In this genre, these production roles are often distributed between scientists, voice actors, and the production team.

The VMA genre has evolved some clear structure and conventions. Hafner (2018) identifies the following seven macro sections: Video Intro; Overview; Researcher's Introduction; Demonstration; Representative Results; Researcher's Conclusion; and Closing Credits. The journal website includes both a written protocol and a video, both of which describe the same procedures. It is interesting to note that the video includes some sections that are not present in the written protocol, namely the Researcher's Introduction and the Researcher's Conclusion. In these sections, the researcher appears on camera and engages with the audience by directly addressing the camera. In the Researcher's Introduction they may introduce themselves and other researchers who appear in the video, forecast the demonstration, explain the significance of the procedure, and invite the audience into the lab. In the Researcher's Conclusion, they may summarize the procedure and any learning, again explain significance, provide tips, wish the viewers well and thank them. Apparently, the adoption of an innovative video genre has legitimated the expression of engaging community identities that are not considered appropriate in the corresponding written genre. While scientists intrude noticeably in these two sections, they remain clearly

distanced in others, like the Overview and Demonstration sections. The VMA therefore provides an especially interesting site to study the multimodal construction of stance and engagement and how this varies across different sections and for different multimodal designers.

4. The multimodal expression of stance and engagement

4.1. Linguistic perspectives

In studies of LSP, there has been considerable interest in interpersonal aspects of communication, including how specialized communicators involve themselves in the discourse to express a stance and how they seek to involve their readers, using discursive strategies of engagement. Broadly conceived, the concept of stance can be defined as the speaker's position vis-à-vis the proposition that they are articulating, vis-à-vis their audience, or vis-à-vis some prior discourse (Jaffe, 2009). An important point to note initially is that, while one can expressly mark stance, a lack of such express marking does not imply a lack of stance. As Jaffe (2009, p. 1) explains, “[a]lthough some forms of speech and writing are more stance saturated than others, there is no such thing as a completely neutral position vis-à-vis one's linguistic productions, because neutrality is itself a stance”. As well as the term “stance”, the terms “metadiscourse” (Hyland, 2005), “evaluation” (Hunston & Thompson, 2000), and “appraisal” (Martin & White, 2005) are also used to refer to similar concepts, studied from a linguistic point of view. For the purposes of this article, I rely on the linguistic taxonomies of stance and engagement that have been developed by Hyland (2005), as an influential framework in the field of LSP.

For Hyland (2005), expressing “stance” means expressing a “textual ‘voice’ or community recognized personality”, including “the ways writers present themselves and convey their judgements, opinions, and commitments” (p. 177). In one sense, this goes beyond the concept of stance referred to above, in that it includes discursive aspects of voice and identity related to the way that authors project themselves in text. In practice, the concept is often operationalized in a narrower way, referring to the position that the speaker adopts towards the proposition. “Stance” conceptualized in this way can in turn be broken down into two main kinds (see also Biber, 2006). First,

“epistemic stance” expresses the speaker’s commitment to the truth value of the proposition, realized by modal auxiliaries like *may* and *must* as well as epistemic markers like *possible*, *likely*, and *certainly*, described in the literature on hedging and boosting (Hyland, 1998, 2005). Secondly, “affective stance” expresses the speaker’s attitude to or emotional evaluation of the proposition, realized by words like *surprisingly* and *understandably*. Hyland (2005) elaborates a number of linguistic categories that are implicated in this kind of stance. They are: hedges and boosters that either reduce or increase a speaker’s commitment to the truth value of a proposition; attitude markers; and self-mention.

Hyland’s concept of “engagement” refers to aspects of the speaker’s position adopted towards the audience, or, more broadly, the way that the presence of the audience is acknowledged. Hyland (2005) defines engagement in the following way:

This is an alignment dimension where writers acknowledge and connect to others, recognizing the presence of their readers, pulling them along with their argument, focusing their attention, acknowledging their uncertainties, including them as discourse participants, and guiding them to interpretations. (p. 177)

In linguistic terms, Hyland identifies reader pronouns, directives, questions, shared knowledge, and personal asides as implicated in engagement strategies. Thus, in specialized texts that involve higher levels of audience engagement, like popular science articles, for example, one would expect to see more of these sorts of features (used in different ways) than in texts involving lower engagement, like research articles, for example.

4.2. Multimodal perspectives

The LSP studies have primarily focused on linguistic features of stance and engagement. At the same time, though, others have examined the way that the use of visuals, writing, speech and all that goes along with it, including gesture, gaze, proxemics and prosody, meaningfully contribute to discourse, arguing that the full complement of multimodal semiotic resources must be considered to understand contemporary communication. For the purposes of this article, there are two related strands of research and theory on multimodal communication that can provide useful insights. First, a social semiotic theory of multimodality, as developed by Kress and his colleagues

(Kress & Jewitt, 2003; Kress, 2010; Kress & van Leeuwen, 2021), provides insight into the way that multimodal artifacts are designed, especially the way that still and moving images make meaning. Second, theories of multimodal interaction illuminate the way that speech and conversation involve the entire body and its co-ordination with other bodies.

The seminal work of Kress and van Leeuwen (2021) provides a detailed theory of visual communication through images. Their model includes interpersonal aspects of communication that relate to the concepts of stance and engagement introduced above. In particular, they talk about “modality” and “interactivity”. “Modality” refers to the extent to which a given image, like a painting or a photograph, is portrayed as a more or less realistic representation. This derives from the linguistic notion of epistemic modality and so can also be thought of as an image producer’s commitment to the “truth” of the image. Kress and van Leeuwen point out that different socio-cultural standards can be applied to judge the notion of what is “realistic” and suggest that the dominant social standard for realism in this respect is one that strives for photorealism. In this sense, images can be manipulated in a range of ways, for example by adjusting colour saturation, so that the quality of the colours appears “washed out” and less photorealistic, or unnaturally saturated and hyperreal. They also address the question of whether scientific diagrams should be considered less realistic, because of the way that they reduce the image to essentialized line drawings and lack photorealism. However, they point out that for the scientific community, such a representation would be judged according to different standards that value an attempt to go “beneath the surface” and distil the true essence of a phenomenon. The kind of image manipulation described here can be used in order to convey elements of both epistemic and affective stance.

“Interactivity” has to do with the way that the position of the viewer is designed. There are three main elements: contact, social distance, and attitude. First, contact between the image producer and the viewer depends on the use of gaze by figures depicted in the image. Where such figures direct their gaze out of the image directly at the viewer, this produces a “demand” image that engages the audience. In contrast, where gaze is not deployed in this way, the result is an “offer” image, a scene “offered” for the viewer’s contemplation. Second, variable social distance between producer and viewer can be achieved through the use of different frame sizes: close shots for intimate/personal relations, medium shots for social relations, and long shots for impersonal relations. Finally, a number of visual resources

contribute to attitude in images. A viewer's subjectivity can be constructed by the use of horizontal camera angle, which can be more direct and frontal (corresponding to greater involvement) or more oblique, angling across the subject (corresponding to greater detachment). Similarly, power relations can be conveyed by vertical camera angle. Where the camera angles up towards the subject, this places the viewer in a lower power relation and when it angles down, it places the viewer in a higher power relation. Viewers can also be provided "objective" images, such as scientific diagrams: when these are presented side on, this represents a "how to" action orientation; when presented top down, this represents a knowledge orientation.

In video genres that include a performance element like the VMAs under examination in this study, stance and engagement can also be expressed through speech as an embodied process involving posture, gesture, facial expression, proxemics and prosody. Useful work here includes research on multimodal interaction conducted from the perspective of conversation analysis and gesture studies. Kendon (2004, p. 7, cited in Abner et al., 2015, p. 438) defines gesture as "visible action when it is used as an utterance or as part of an utterance", a definition that encompasses gesturing with arms and hands but also includes actions taken by other parts of the body, like nods of the head, shrugs, and facial movements. Gestures can be divided into those that are interactive and used to manage conversation with an interlocutor, and those that are representational and communicate something about the content of the utterance. These latter can further be divided into deictic gestures, i.e., pointing in space, iconic gestures that depict concrete imagery, metaphorical gestures depicting abstract imagery and conventional gestures like a "thumbs up" or "thumbs down". Both interactive and representational gestures could function to express a position towards what one is saying and towards one's audience, i.e., to express stance and engagement. Furthermore, work on multimodal interaction that has explicitly considered the notion of stance, has shown how not only gesture but also proxemics, facial expression, posture, and body position contribute to the construction of an affective stance, allowing speakers to align with and disalign with others (Goodwin et al., 2012). Much of the research conducted in this tradition of multimodal interaction analysis assumes a context of face-to-face interaction. In contrast, the present study attempts to shed light on the way that such multimodal semiotic resources contribute to meaning making on the screen for a mediated, video genre.

4.3. Visual communication in science

Existing multimodal work on academic performance genres focuses mainly on conference presentations, including TED Talks, examining the way that academic presenters make use of different kinds of visuals as an important part of their performance. Rowley-Jolivet (2002, 2004) examines the use of slides in conference presentations in geography and medicine and develops a typology of visuals, including Figurative visuals, Graphical visuals, Numerical visuals, and Scriptural visuals. Figurative visuals are photographs or artistic representations and include three subtypes (the third suggested by Xia, 2023a). Figurative I images consist of “ordinary photographs”; Figurative II images consist of photographs that have been enhanced, for example with annotation, to highlight a single feature of the represented object; and Figurative III images are “drawings and caricatures” that “artistically represent [scientific] phenomena” in sometimes playful ways (Xia, 2023a, p. 77). Graphical visuals refer to graphs, diagrams, and maps. Numerical visuals are comprised of mathematical formulae and associated calculations. Finally, Scriptural visuals are those that make use of writing. This empirical work provides a useful starting point when considering stance and engagement in the analysis of visual resources in video genres as the selection of a particular kind of visual could have a particular effect on the kind of “textual voice” or stance that is conveyed as well as be designed to meet the expectations of, and so engage with, the disciplinary community.

5. The study

5.1. Aims and approach

This study aims to determine how stance and engagement is multimodally co-constructed in video methods articles (VMAs). More particularly, the research questions are:

RQ1: How is epistemic and affective stance co-constructed in different sections of VMAs, paying particular attention to the way that multiple semiotic resources are used to express a “textual voice” and a presentation of self?

RQ2: How is engagement co-constructed in VMAs, paying particular attention to the way that multiple semiotic resources are used to align with and seek to involve the audience?

The study builds on an earlier, multi-perspective genre analysis of the VMA (Hafner, 2018), which established understandings of core features of the genre, including its intertextual communicative context, purpose, audience, move structure and multimodal realizations, development over time, and aspects of the collaborative construction process, as described above. Here, I extend that work by making use of a qualitative approach to multimodal discourse analysis, theoretically grounded in a social semiotic approach to multimodality (see above, e.g., Kress, 2010). A qualitative approach is common in this kind of multimodal discourse analysis because it allows the analyst to provide a rich account of the semiotic features observed, including precisely how varied semiotic resources contribute to meaning making. Such qualitative inquiry is also common in exploratory studies that seek to build or expand existing theory, as the present one does.

5.2. Data sources

This study draws on a total of 11 VMAs published by the *Journal of Visualized Experiments* (*JoVE*) between its inception in 2006 and 2016 (one VMA per year from the time the journal was launched until the time of data collection for the initial project). In order to limit the potential effect of discipline on observations, only VMAs from biology were collected. Biology VMAs with the highest number of citations in Scopus at the time were collected, as such VMAs can be said to have received a level of community endorsement and so likely be examples of best practice. The videos address a range of topics in biology, including behavioural tests of mice, a surgical technique for mice, and various lab procedures, for example detecting proteins or isolating exosomes. The larger project also relied on a number of other data sources, to provide multiple perspectives on the genre. These include: extracts from the *JoVE* website, such as Author Guidelines and About Us pages; comments on articles in the comments section of the webpage (articles had received between zero and 71 comments per article at the time of data collection); interviews with specialist informants, including the founding CEO of *JoVE* and a biologist end-user of the videos.

While relying on insights from this larger data set, the present study focuses analysis on the videos themselves. It provides both a general picture of the way that stance and engagement work in different moves of the VMA, as well as a more specific analysis of selected moves in selected VMAs. Based on the earlier genre analysis (see above), it was clear that an analysis of the Overview section, which provides a summary of the article, would provide

an opportunity to observe a more “neutral” scientific stance being enacted, while an analysis of the Researcher’s Introduction and Researcher’s Conclusion would allow observation of a more involved stance along with likely engagement strategies. An initial analysis allowed for a purposive selection of three contrasting VMAs: one considered high engagement (VMA4), one considered medium engagement (VMA6), and one considered low engagement (VMA8). Because of space constraints, the data presented in this article are from VMA4 and VMA8 only.

5.3. Analytical procedures

All videos were initially viewed multiple times and linguistically transcribed. All of the sections present in the videos were examined to identify typical scientist and audience involvement. The selected sections (Overview, Researcher’s Introduction, and Researcher’s Conclusion) of all videos were reviewed and initial notes on multimodal expression were made. This led to the selection of the three high, medium, and low engagement videos. The Overview, Researcher’s Introduction, and Researcher’s Conclusion of these three videos were further multimodally transcribed, using the *ELAN* software (Max Planck Institute for Psycholinguistics, 2023), which allows the user to annotate multiple levels, corresponding to different kinds of semiotic choices. The transcription method draws on the framework developed by Baldry and Thibault (2006) and by Hafner (2014). First, relevant sections of the video were annotated to identify individual shots, which make for a robust unit of analysis. Where a single shot included multiple visual frames, as where a pan or zoom was involved or where an animated diagram introduced a new visual configuration, these visual frames were further segmented and annotated. For each such unit identified, the following features were transcribed: 1) speech; 2) action; 3) gesture (including facial expression); 4) gaze; 5) soundtrack; 6) camera position and perspective (vertical angle, horizontal angle, frame, i.e., close-up, medium, long shots); 7) text on screen. This detailed transcription process is an effective way of making the researcher aware of the way that different semiotic resources are used (whether consciously or not) and how they relate as part of a multimodal ensemble.

6. Findings

6.1. Overview of scientist and audience involvement by section

Analysis shows that different sections of the VMA afford different opportunities for scientist and audience involvement, with a range of different possible discourse roles (textual “voice”) for the scientist authors. Typical realizations of these roles are summarized in Table 1.

	Typical scientist role, i.e., “textual voice”	Typical visual participant(s)	Typical aural participant(s)	Typical setting
Video Intro	Named author	JoVE logo with animation, text on screen	JoVE jingle, background music	NA
Overview	Scientific expert, lab demonstrator	Animated diagrams, equipment, scientist(s), JoVE logo	Voice actor, background music	Lab, where applicable
Researcher’s Introduction	Presenter host, lab demonstrator, scientific expert	Scientist, text on screen (e.g., lower thirds), JoVE logo	Scientist, background music	Lab, office, classroom, university campus
Demonstration	Lab demonstrator	Text on screen, scientist(s), equipment, JoVE logo	Voice actor, background music	Lab
Representative Results	Scientific expert	Graphs, charts, data images, JoVE logo	Voice actor, background music	NA
Researcher’s Conclusion	Presenter host, lab demonstrator, scientific expert	Scientist(s), JoVE logo	Scientist, background music	Lab, office, classroom
Closing Credits	NA	Text on screen, JoVE logo, Production team as named contributors	Background music	NA

Table 1. Typical realizations of scientist roles/textual voices

When it comes to involvement of the scientist, this can be achieved through visual representation, aural participation (i.e., speech), and linguistic means. The most obvious scientist involvement is in the Researcher’s Introduction and Researcher’s Conclusion sections, where the scientist appears facing the camera and speaking into the camera. These are also the two sections with the greatest audience involvement as the scientist directly addresses the audience, making use of their visual presence, their own voice, and linguistic expressions of engagement. In these two sections, the scientist plays the role of a presenter host, welcoming viewers, either previewing or summarizing

the innovative methods showcased, and often closing by thanking them or wishing them well. I will go into this in more detail below. In contrast, other sections of the VMA involve the scientist and audience to a lesser extent. These other sections are usually narrated by a professional voice actor, thereby removing the scientist voice from the audio. Where the scientist appears in the visuals physically, they play the role of a lab demonstrator, who operates equipment and does not visually engage with the audience. The voice-over, realized by the voice actor, explains goals and significance of the procedures (in the Overview), provides instructions on procedural steps often using imperatives (in the Demonstration), and describes expected results (in the Representative Results). A much lower degree of scientist and audience involvement is observed in these other sections. In the following sections, I limit observations to the Overview, Researcher's Introduction and Researcher's Conclusion sections, examining the semiotic resources that are used to co-construct stance and engagement. As noted above, I compare VMA4 (high engagement, referred to as VMA4-HE) with VMA8 (low engagement, referred to as VMA8-LE).

6.2. The Overview: The stance of the scientific expert and lab demonstrator

In the Overview section, one stance enacted is that of a conceptual scientific expert. The scientist is projected in the text in only a minimal way, as a result of a range of semiotic choices: visually, there is heavy reliance on animated diagrams and shots of lab procedures; aurally, the scientist is absent, with the scripted voice-over read by a professional voice actor; linguistically, markers of stance and engagement are largely missing and the text is characterized by frequent nominalization and passivization, removing the scientist agents from the text.

In VMA8-LE, there is no trace of scientist involvement at all: the Overview relies entirely on animated diagrams (graphical visuals), with labelled 2D representations of test tubes illustrating the experimental procedures (combining different solutions) and depicting resulting chemical processes in abstract, conceptual ways, visually separating the procedures from the material world of the lab (see Figure 1). Explicit linguistic stance and engagement markers are entirely absent from the voice-over. The effect of these semiotic choices is to construct a scientific stance that prioritizes conceptual representations of materials and procedures over human agents and their material world. As already noted, such a stance could meet

disciplinary expectations in a way that is perceived as expert and authoritative, by presenting a conceptual “essence” of the experiment (Kress & van Leeuwen, 2021).




		
1 min 3.1 seconds (5a)	1 min 4.4 seconds (5b)	1 min 5.8 seconds (5c)
[The final step] is to execute a T X T L	reaction	using the reagents just made

Figure 1. Animated diagram with voice-over (Overview, VMA8-LE, shots 5a-5c)¹

VMA4-HE presents a similar kind of scientific expert stance, with a largely author-evacuated text. Here though, slight traces of scientist involvement are visible. That Overview begins with the phrase, “In this video we will demonstrate the power of...”, one single instance of self-mention. In addition to the kind of graphical visuals used in VMA8-LE, the section also relies on a considerable number of figurative visuals, i.e., images of the lab setting that show the scientist’s body (partially), equipment, and materials. In this sense, the scientist is involved, projected in the text—a kind of visual self-mention. However, this involvement is very limited. The images are usually shot from behind and feature only hands holding equipment (see Figure 2). The scientist body is represented in an impersonal manner as one component of the laboratory equipment, like the machine that drives the experiment. This is not a social body: the scientist never engages with the audience, for example, turning to the camera and explaining what she is doing. Indeed, as already noted, the scientist never speaks. This is the same use of semiotic resources that will be picked up later in the Demonstration section, though more of the scientist will become visible—as well as hands, we will see arms and torso and on rare occasions the full body. These representations index what I have referred to as the lab demonstrator stance (Table 1).

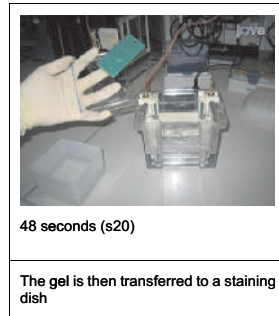


Figure 2. Limited scientist involvement (Overview, VMA4-HE, shot 20)¹

6.3. The Researcher's Introduction and Conclusion: The stance of the presenter host

As noted earlier, the Researcher's Introduction (RI) and Researcher's Conclusion (RC) sections provide greater opportunities for scientist involvement and audience engagement. Specifically, there is an opportunity to adopt the stance of a presenter host, for example introducing the video to the audience. In the RI for VMA4-HE, this is accomplished by adopting four out of the seven genre moves available (see Hafner, 2018): the scientist begins by introducing herself with her affiliation ("Hello, *I* am... from the laboratory of..."); then, she forecasts the demonstration ("Today, *I* will draw *your* attention to a widely non-famous procedure"); next, she explains its significance ("We use Kang's staining protocol in *our* laboratory for nearly every gel-based application"); finally, she closes by inviting the audience ("So *let's* get started!"). As can be seen from the examples, the language used includes multiple stance and engagement markers, which help to establish the presence of the scientist and acknowledge the presence of the audience. This is reinforced by a range of semiotic choices. First, the scientist appears on the video, at medium close distance, facing the camera directly with a level vertical camera angle. Her gaze is directed straight at the camera producing an engaging demand image. She begins her speech with a visible smile which fades to a more neutral expression as she continues to explain but returns at the end of the video.

An analysis of the final invitation to the audience shows how it is accomplished through a complex orchestration of multiple semiotic resources, including the use of language, prosodic elements of speech, gaze,

gesture, facial expression, and the body. Over the duration of this 3.8 second shot, the scientist engages the audience with a wide smile, conveying happy, excited attitudinal stance. As in the rest of the RI, a direct, level camera angle is used at medium distance with the scientist gaze on the camera creating a demand image. As she articulates her line, “So let’s get started” she gestures, first nodding (on “let’s”, Figure 3, first frame) then raising both hands into the frame in a thumbs up position, bringing the hands across her body from her left to her right (on “started”, Figure 3, second frame), in a deictic gesture that signals her intended direction of travel. The intonation of “started” reinforces her excitement with noticeably high pitch on “start” that then drops dramatically. Finally, she turns her body to the right and brings her hands down with two index fingers outstretched in front of her again making a deictic gesture (Figure 3, third frame), before moving out of the frame to her right. In this sequence, the use of the body—pointing and turning—serves to issue the scientist’s invitation as one to join her in a specific place (the lab, one assumes), off camera to her right, making the verbal utterance “let’s get started” more concrete by going beyond the words and adding a spatial meaning. Similarly, the attitude of happy excitement that is conveyed is signalled by the use of face and prosody, rather than through the words on their own.

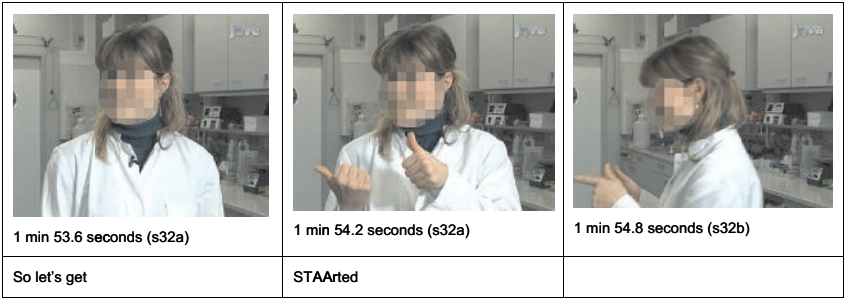


Figure 3. Multimodal engagement in the higher engagement VMA (Researcher’s Introduction, VMA4-HE, shot 32)¹

The lower engagement VMA8-LE provides an interesting contrast. Here, the RI consists of only two of the seven available genre moves: first, explaining the significance of the procedure (“This method can help test circuits in the field of synthetic biology”); and second, introducing a team member (“Assisting in the procedure will be Clare Hayes, a research assistant in our group”). We see some scientist presence in this video, especially as they are

physically present on camera, the lead author is speaking, and employs self-mention, “our group”. However, linguistically there is little acknowledgment of the audience and none from the lead author at all. The visual composition is a similar direct shot but the frame is tighter and the camera closer. At the same time, the lead author directs his gaze slightly to the side of the camera on his right, rather than directly at it, thereby lessening the engaging effect. He maintains a neutral facial expression throughout (see Figure 4). When he introduces the group’s research assistant, the visual composition becomes a bit more engaging: she is shown from the side sitting in front of her computer and looking at the screen; she then turns to face the camera, gazes into it and smiles (see Figure 4).




		
1 min 26.7 seconds (s7)	1 min 34.7 seconds (s8a)	1 min 36.5 seconds (s8b)
... The implications of this technique extend towards increasing the speed of synthetic biological design by removing the need to conduct all prototyping steps in vivo.	Assisting in the procedure will be Clare Hayes,	a research assistant in our group.

Figure 4. Multimodal engagement in the lower engagement VMA (Researcher’s Introduction, VMA8-LE, shots 7-8)

7. Discussion and conclusions

While this exploratory study is limited to a relatively small number of video texts, it nevertheless provides interesting insights into the way that stance and engagement are multimodally co-constructed in the VMA genre. What it means to take a multimodal approach to stance and engagement is to consider how the full range of multimodal semiotic resources contribute to the realization of author and audience involvement in the multimodal text. For example, drawing on existing conceptualizations of stance and engagement, the analyst must consider how the scientist author intrudes into the text to express a textual voice or acknowledges the presence of the audience through a range of semiotic resources. The findings identify three

main textual voices expressed multimodally: the scientific expert, which is expressed in the Overview section through animated diagrams and heavy reliance on linguistic features of a highly nominalized scientific register with very few stance and engagement markers; the lab demonstrator, which is visually expressed in the Overview through the scientist's presence in the lab; and the presenter host, expressed in the Researcher's Introduction and Researcher's Conclusion through use of engaging visuals of the scientist and use of stance and engagement markers. These multiple textual voices speak to the hybridity of the genre, which appears to combine elements of more popular media genres with elements of more specialized scientific ones. All of these voices are realized through multimodal orchestrations of semiotic resources.

The findings point to more explicit expression of stance and engagement in the Researcher's Introduction (RI) and Researcher's Conclusion (RC) sections, where the scientist is not only visually present on screen but also speaking. This contrasts with other sections, where a voice-over is typically provided by a professional voice actor, reducing (but not removing) scientist involvement. For example, in the Overview, scientist presence is visually minimized, and the focus of the video text is on conceptual processes expressed through graphical visuals, as well as lab procedures expressed through figurative visuals of the scientist, equipment, and materials. This focus echoes the tendency of academic writing in science to downplay the researcher's role in scientific discovery. The almost total lack of a scientist presence itself serves to establish a stance that portrays the procedures as neutral and objective. The Overview also tends to lack explicit engagement features. There are no engaging glances at the camera, nods to the audience, or asides mentioned as the scientist goes about demonstrating the lab work, which would certainly feature in similar entertainment genres demonstrating procedures, like cooking shows for example. By comparison, we sometimes see plentiful attempts at engagement in RI and RC sections. These two new sections represent innovations that echo the approach taken in popular science articles, to focus on researchers rather than research procedures (Myers, 1990). In addition, the findings as they relate to RI and RC sections suggest considerable individual variation in terms of researcher and audience involvement, which varied in the two cases considered.

This study extends previous work on VMAs by paying closer attention to the representation of the scientist body, including attention to not only dress, gesture, facial expression, and prosody, but also the emplacement of the

body in (typically) a lab or office setting and even filmic choices such as the framing of the body with close-up or medium shots, and whether the focus allows the viewer to see the scientist's face from the front or is limited to oblique shots of hands, arms and torso, clad in lab coat and gloves as they go about performing the experiment. The findings show how these resources, including movement through space, serve to add new dimensions of meaning in relation to stance (especially attitudinal stance) and engagement. It is also interesting to consider the way that the body in such video genres is co-constructed by the scientist and the production team in order to construct stance and engagement. The former produces the performance, while the latter directs, films, and edits it, presumably contributing to on-site choices of location, stage directions such as whether to sit or stand, and the use of particular filmic resources like framing, discussed above. More research is needed in order to better understand this process, with studies that draw on ethnographic techniques like observation and interview to focus attention on the process of video creation, in addition to the product. A study of process would also shed light on the extent to which scientists contribute to decisions related to video production, like the generation of animations, editing choices and so on, in the way that we know they do in TED Talks (Xia, 2023b). In addition, it would be interesting to see how much preparation is involved prior to the performance, including to what extent the use of the body is directed and rehearsed.

There seem to be some generic constraints on the way that the body is represented in this genre as suggested by differences observed in Overview, RI and RC sections. It is interesting to compare the findings of this study with work on the role of the body in TED talks, another video performance genre. Taking an enactive-ecological approach, Harrison's (2021) case study of the science communicator Brian Cox highlighted an interplay between speaker, gesture, slide and audience, at a moment in which the speaker directed gaze between his own gesturing hands, which were in turn aligned with information on his slides, and the audience in such a way as to convey his own intense engagement with the subject matter and involve the audience. In contrast, the use of gesture in the VMAs studied here appears more limited: for the most part, speakers expressing a presenter host stance direct their gaze either towards or directly into the camera but are framed in such a way that hands are often outside the shot and available gesture is largely limited to facial expressions and nods of the head. Visually, the body here resembles that of a TV presenter on the news or a participant in a media interview, and this choice

of framing could well be a generic choice of the media production team, that contributes to the construction of this hybrid media/scientific genre. By contrast, in TED talks, the presenter can be framed from multiple distances: from close-up to long shots (Xia and Hafner, 2021; Xia, 2023a). In other words, these two different video performance genres seem to differ in the conventional way that meaning is constructed through the presenter body. Of course, there are other obvious differences as well, including the fact that the TED Talk is a recording of a (highly structured and carefully prepared) live talk. Similar observations are made by Bernad-Mechó and Valeiras-Jurado (2023) who found that, compared to a TED talk, a PBS YouTube video relied far more on filmic resources (such as cuts) than embodied resources, as a way to engage the audience.

This study contributes to a growing body of research in LSP that examines multimodal expression in specialized genres. In this work, it has been necessary to extend existing frameworks of genre analysis to account for multimodal expression in video and social media genres. The analytical framework adopted in this study drew on insights into stance and engagement from the perspective of linguistics (Hyland, 2005), visual design (Rowley-Jolivet, 2002, 2004; Kress & van Leeuwen, 2021; Xia, 2023a), and embodied interaction (Goodwin et al., 2012). The analysis demonstrated that it is possible to examine stance and engagement in video-based performance genres by taking into account all of these semiotic resources. In particular, the framework extended previous work by properly accounting for embodied meanings expressed in gesture, facial expression, and gaze. Recent studies in LSP, like that of Jaworska (2023) and Xia (2023a), also attempt this kind of multimodal analysis. However, in these studies less detailed attention is paid to the role of the body in meaning making. The framework adopted in this study offers some suggestion of how this additional dimension can be accounted for in multimodal analysis for LSP (see also Bernad-Mechó & Valeiras-Jurado, 2023).

Specifically, Hyland's concepts of stance and engagement were extended by taking into account additional (filmic and embodied) resources that were found to index "alignment", which, drawing on Goodwin et al. (2012), I have found to be a helpful notion in understanding the way that the body indexes an affective stance and engagement with the audience. This expansion complements recent empirical work on social media like that of Luzón (2023) which also adopts a multimodal approach to stance and engagement, expanding on Hyland's linguistic categories by integrating visual elements such as emojis, memes and images. Luzón operationalizes these elements as

stance and engagement markers, allowing for a quantitative treatment of the use of visual and linguistic markers in academic tweets. In contrast, this case study of VMAs orients to the way that multiple semiotic resources are orchestrated in the performance of textual voices and alignments and disalignments with the audience.

The study suggests that multimodal resources can contribute to creative, intertextual appropriation of semiotic resources (Bhatia, 2017). Such intertextual appropriations can be happening at multiple levels. We see here appropriations: 1) at the level of text, with the use of multimodal resources indexing stance and engagement; 2) at the level of genre, with innovative features like the addition of engaging Researcher's Introduction and Researcher's Conclusion sections; and, 3) at the level of disciplinary culture and professional identity, with the performance of both more popular and more academic identities. The study also suggests that such appropriation can occur as a result of collaborative construction: it might not be a coincidence that the kind of hybridity that we see here emerges in a genre that is co-constructed by media professionals and scientists. Such an interpretation would find support in Rowley-Jolivet and Carter 'Thomas' (2019) work on podcasts that showed how scientists and journalists work together to produce a kind of hybrid discourse, with journalists tending to make use of more popular strategies such as the use of puns, and scientists making use of more academic strategies.

In conclusion, it is to be hoped that this study also provides useful food for thought for teachers of English for science students who are interested in engaging their students with emerging performance genres in video, as a form of digital multimodal composing (Hafner & Miller, 2019). In particular, the findings could shed light on how students might adopt a range of textual voices, making use of multiple modes to present themselves in more or less involved and engaging ways, depending on their communicative purpose and their intended audience. They could also shed light on possible generic elements of the text construction with respect to multimodal orchestration, for example, the way that certain kinds of multimodal configurations may be expected to perform certain functions. In LSP research and pedagogy, there is increasing interest in multimodal practices. With the VMA, we have a genre with a clearly distributed production format that is the result of collaborative efforts of a range of professionals. Future research on this and other genres would benefit from more reflection on the distributed nature of composing. In particular, the way that laboratory teams work together to produce

academic texts, along with their multimodal elements like charts, graphs, and figures, could provide an interesting focus for future research that leads to a better understanding of the discourse competence of academic writers.

Article history:

Received 21 July 2023

Received in revised form 3 October 2023

Accepted 10 October 2023

Acknowledgements

I would like to thank the editors for their support and encouragement with this article, as well as the anonymous reviewers who provided useful feedback on the first draft of this article. In addition, I also feel indebted to Simon Harrison for the many interesting conversations that we had about multimodal approaches to communication and interaction, before, during and after the writing of this paper.

References

- Abner, N., Cooperrider, K., & Goldin-Meadow, S. (2015). Gesture for linguists: A handy primer. *Language and Linguistics Compass*, 9(11), 437–451. <https://doi.org/10.1111/lnc3.12168>
- Baldry, A., & Thibault, P. J. (2006). *Multimodal transcription and text analysis*. Equinox.
- Bernad-Mechó, E., & Valeiras-Jurado, J. (2023). Multimodal engagement strategies in science dissemination: A case study of TED talks and YouTube science videos. *Discourse Studies*, 14614456231161755. <https://doi.org/10.1177/14614456231161755>
- Bhatia, V. K. (2017). *Critical genre analysis: Investigating interdiscursive performance in professional practice*. Routledge.
- Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, 5(2), 97–116. <https://doi.org/10.1016/j.jeap.2006.05.001>
- Breeze, R. (2019). Continuity and change: Negotiating relationships in traditional and online peer review genres. In M. J. Luzón & C. Pérez-Llantada (Eds.), *Science communication on the internet* (pp. 107–129). John Benjamins.
- Carter-Thomas, S., & Rowley-Jolivet, E. (2017). Open science notebooks: New insights, new affordances. *Journal of Pragmatics*, 116, 64–76. <https://doi.org/10.1016/j.pragma.2016.12.003>
- Goffman, E. (1981). *Forms of talk*. University of Philadelphia Press.
- Goodwin, M. H., Cekaite, A., & Goodwin, C. (2012). Emotion as stance. In A. Perakyla & M.-L. Sorjonen (Eds.), *Emotion in Interaction* (pp. 16–41). Oxford University Press.
- Hafner, C. A. (2014). Embedding digital literacies in English Language Teaching: Students' digital video projects as multimodal ensembles. *TESOL Quarterly*, 48(4), 655–685. <https://doi.org/10.1002/tesq.138>
- Hafner, C. A. (2018). Genre innovation and multimodal expression in scholarly communication: Video methods articles in experimental biology. *Ibérica, Journal of the European Association of Languages for Specific Purposes*, 36, 15–41. Retrieved from <https://www.revistaiberica.org/index.php/iberica/article/view/121>
- Hafner, C. A., Harrison, S., Ho, W. Y. J., & Kwan, B. S. C. (2023). Digital mediation in ESP genres. *English for Specific Purposes*, 71, 115–122. <https://doi.org/10.1016/j.esp.2023.03.002>
- Hafner, C. A., & Miller, L. (2019). *English in the*

disciplines: A multidimensional model for ESP course design. Routledge.

Hafner, C. A., & Pun, J. (2020). Editorial: Introduction to this special issue: English for academic and professional purposes in the digital era. *RELC Journal*, 51(1), 3–13. <https://doi.org/10.1177/0033688220917472>

Harrison, S. (2021). Showing as sense-making in oral presentations: The speech-gesture-slide interplay in TED talks by Professor Brian Cox. *Journal of English for Academic Purposes*, 53, 101002. <https://doi.org/10.1016/j.jeap.2021.101002>

Hunston, S., & Thompson, G. (Eds.). (2000). *Evaluation in text: Authorial stance and the construction of discourse*. Oxford University Press.

Hyland, K. (1998). *Hedging in scientific research articles*. John Benjamins.

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173–192. <https://doi.org/10.1177/1461445605050365>

Hyland, K., & Zou, H. J. (2022). Pithy Persuasion: Engagement in 3 Minute Thesis Presentations. *Applied Linguistics*, 43(1), 21–44. <https://doi.org/10.1093/applin/amab017>

Jaffe, A. (Ed.). (2009). *Stance: Sociolinguistic perspectives*. Oxford University Press.

Jaworska, S. (2023). But then something happened: A critical multimodal genre analysis of corporate image repair videos. *English for Specific Purposes*, 69, 95–108. <https://doi.org/10.1016/j.esp.2022.10.004>

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kress, G., & Jewitt, C. (2003). Introduction. In C. Jewitt & G. Kress (Eds.), *Multimodal literacy* (pp. 1–18). Peter Lang.

Kress, G., & van Leeuwen, T. (2021). *Reading images: The grammar of visual design* (Third edition). Routledge.

Kuteeva, M., & Mauranen, A. (2018). Digital academic discourse: Texts and contexts: Introduction. *Discourse, Context & Media*, 24, 1–7. <https://doi.org/10.1016/j.dcm.2018.06.001>

Luzón, M. J. (2013). Public communication of science in blogs: Recontextualizing scientific discourse for a diversified audience. *Written*

Communication, 30(4), 428–457. <https://doi.org/10.1177/0741088313493610>

Luzón, M. J. (2023). Multimodal practices of research groups in Twitter: An analysis of stance and engagement. *English for Specific Purposes*, 70, 17–32. <https://doi.org/10.1016/j.esp.2022.10.006>

Luzón, M. J., & Pérez-Llantada, C. (2019) (Eds.). *Science communication on the internet*. John Benjamins.

Luzón, M. J., & Albergo-Posac, S. (2020). ‘Had a lovely week at #conference2018’: An analysis of interaction through conference tweets. *RELC Journal*, 51(1), 33–51. <https://doi.org/10.1177/0033688219896862>

Luzón, M. J., & Pérez-Llantada, C. (2022). Digital genres in academic knowledge production and communication: Perspectives and practices. In *Digital Genres in Academic Knowledge Production and Communication*. Multilingual Matters. <https://doi.org/10.21832/9781788924726>

Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan.

Max Planck Institute for Psycholinguistics. (2023). *ELAN* [Computer software]. <https://archive.mpi.nl/tla/elan>

McGrath, L. (2016). Open-access writing: An investigation into the online drafting and revision of a research article in pure mathematics. *English for Specific Purposes*, 43, 25–36. <https://doi.org/10.1016/j.esp.2016.02.003>

Myers, G. (1990). *Writing biology: Texts in the social construction of scientific knowledge*. University of Wisconsin Press.

Pérez-Llantada, C. (2016). How is the digital medium shaping research genres? Some cross-disciplinary trends. *ESP Today*, 4(1), 22–42.

Reid, G., & Anson, C. M. (2019). Public- and expert-facing communication: A case study of polycontextuality and context-collapse in Internet-mediated citizen science. In M.-J. Luzón & C. Pérez-Llantada (Eds.), *Science communication on the internet* (pp. 219–238). John Benjamins.

Rowley-Jolivet, E. (2002). Visual discourse in scientific conference papers: A genre-based study. *English for Specific Purposes*, 21(1), 19–40. [https://doi.org/10.1016/S0889-4906\(00\)00024-7](https://doi.org/10.1016/S0889-4906(00)00024-7)

Rowley-Jolivet, E. (2004). Different visions, different visuals: A social semiotic analysis of field-specific visual composition in scientific conference presentations. *Visual Communication*, 3(2), 145–175. <https://doi.org/10.1177/147035704043038>

- Rowley-Jolivet, E., & Carter-Thomas, S. (2019). Scholarly soundbites: Audiovisual innovations in digital science and their implications for genre evolution. In M. J. Luzón & C. Pérez-Llantada (Eds.), *Science communication on the internet* (pp. 81–106). John Benjamins.
- Sancho Guinda, C. (2015). Genres on the move: Currency and erosion of the genre moves construct. *Journal of English for Academic Purposes*, 19, 73–87. <https://doi.org/10.1016/j.jeap.2015.07.001>
- Scotto di Carlo, G. (2014). The role of proximity in online popularizations: The case of TED talks. *Discourse Studies*, 16(5), 591–606. <https://doi.org/10.1177/1461445614538565>
- Tardy, C. M. (2023). How epidemiologists exploit the emerging genres of Twitter for public engagement. *English for Specific Purposes*, 70, 4–16. <https://doi.org/10.1016/j.esp.2022.10.005>
- Xia, S. (2023a). Explaining science to the non-specialist online audience: A multimodal genre analysis of TED talk videos. *English for Specific Purposes*, 70, 70–85. <https://doi.org/10.1016/j.esp.2022.11.007>
- Xia, S. (2023b). *Popularizing Science in the Digital Era: A Multimodal Genre Perspective on TED Talk Videos*. Routledge.
- Xia, S., & Hafner, C. A. (2021). Engaging the online audience in the digital era: A Multimodal analysis of engagement strategies in TED talk videos. *Ibérica, Journal of the European Association of Languages for Specific Purposes*, 42, 33–57. <https://doi.org/10.17398/2340-2784.42.33>
- Zou, H., & Hyland, K. (2020). Academic blogging: Scholars' views on interacting with readers. *Ibérica, Journal of the European Association of Languages for Specific Purposes*, 39, 267–294. <https://doi.org/10.17398/2340-2784.39.267>

Christoph A. Hafner is Associate Professor in the Department of English, City University of Hong Kong and President of the Asia-Pacific LSP & Professional Communication Association. He has published widely on English for specific purposes, digital literacies, and language learning and technology. His latest book is *Understanding Digital Literacies: A Practical Introduction* (2nd edition) (Routledge, 2021, co-authored with Rodney Jones).

NOTES

¹ Images used with permission

Appendix: VMAs referred to in analysis

- VMA4 Dyballa, N., & Metzger, S. (2009). Fast and Sensitive Colloidal Coomassie G-250 Staining for Proteins in Polyacrylamide Gels. *JoVE (Journal of Visualised Experiments)*, (30), e1431–e1431. <https://doi.org/10.3791/1431>
- VMA8 Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., & Noireaux, V. (2013). Protocols for Implementing an Escherichia coli Based TX-TL Cell-Free Expression System for Synthetic Biology. *JoVE (Journal of Visualised Experiments)*, (79), e50762–e50762. <https://doi.org/10.3791/50762>