Engaging the online audience in the digital era: A multimodal analysis of engagement strategies in TED talk videos

Sichen Ada Xia & Christoph A. Hafner
The Chinese University of Hong Kong, Shenzhen & City University of Hong Kong
xiasichen@cuhk.edu.cn & c.hafner@cityu.edu.hk

Abstract

Scientific popularization genres have attracted the attention of researchers in esp. Authors of such genres have been found to employ a variety of strategies to engage the readers. However, most studies of engagement in scientific popularization tend to focus on written genres. For this reason, they have mostly examined only the linguistic mode and rarely gone beyond this to take multimodal resources into consideration. The present study investigates how multimodal semiotic resources are combined in digital media to engage the online audience in an influential online scientific popularization genre, that is, TED talks. Drawing on a corpus of 28 highly viewed TED talks from the area of biology, this study provides an in-depth analysis of multimodal engagement strategies in one of those talks. The analysis shows that the engagement of online viewers in a digitally mediated scientific popularization genre like TED talks tends to be achieved by the combination of multiple semiotic resources afforded by digital technologies, such as speech, gestures, gaze, visuals, distance of shot, perspective of shot and angle of shot. The study also reveals five strategic multimodal configurations employed to achieve viewer engagement, with the focus variously on engagement through visual aids, a long shot, gaze, questions and reference to personal emotions.

Keywords: digital genres, multimodality, scientific popularization, engagement.

Resumen

Implicando a la audiencia en línea en la era digital: un análisis multimodal de las estrategias de implicación en vídeos de charlas de TED

Los géneros de divulgación científica han captado la atención de los investigadores del ámbito del inglés con fines específicos. Se ha constatado que los autores de tales géneros emplean diferentes estrategias para involucrar al lector. Sin embargo, la mayoría de los estudios sobre la noción de *implicación* (*engagement*, en la bibliografía anglófona) en la divulgación científica se han centrado en los géneros escritos. Por ello, hasta el momento se ha analizado fundamentalmente el modo lingüístico y apenas se han explorado otras áreas como los recursos multimodales. El presente artículo analiza el modo en que se combinan los recursos semióticos multimodales en los medios digitales para implicar a la audiencia en línea en un influyente género de divulgación científica en Internet: las charlas de TED. El análisis llevado a cabo muestra que es una combinación de múltiples recursos semióticos posibilitados por las tecnologías digitales aquello que permite implicar a la audiencia en línea en un género de divulgación científica digital como las charlas de TED. Entre esos recursos se encuentran el propio discurso, los gestos, la mirada, los elementos visuales o la distancia, la perspectiva y el ángulo de la toma. Este trabajo también formula cinco configuraciones multimodales empleadas estratégicamente para involucrar a la audiencia, con especial atención al uso de elementos visuales auxiliares, planos largos, miradas, preguntas y referencias a las emociones personales.

Palabras clave: géneros digitales, multimodalidad, divulgación científica, implicación.

1. Introduction

Scientific popularization, namely, the dissemination of scientific knowledge to a lay audience, has been extensively discussed by researchers. A testament to this is the continuous definitional debate in which the popularizing process has been variously recognized as a process of vulgarization, translation and most recently, recontextualization (Bondi, Cacchiani & Mazzi, 2015; Calsamiglia & Ferrero, 2003; Gotti, 2014; Luzón, 2013; Myers, 2003). This pejorative-towards-neutral perception of popularization reflects the change in the understanding of the ownership of scientific knowledge. While science used to be regarded as exclusively belonging to specialists due to its "progressive specialization and ever-increasing 'technicity'" (Calsamiglia, 2003: 140), more recently the general public has been considered to be eligible to learn about it. One possible reason for this is that a good public understanding of science can engender multiple benefits. Thomas and Durant (1987) argue that popularization benefits multiple interests including the scientific enterprise itself, national economies,

national power and influence, individuals and those whose work could be intellectually and aesthetically inspired by science. Furthermore, in the present time, there is a widespread trend in academia to encourage academics to achieve impact on society in addition to publishing groundbreaking research articles. This means that from the perspective of scientists, it is increasingly important for them to harness the skills of popularizing science.

The existing literature suggests two possible models of scientific popularization. In the first model, media serves as a bridge between scientists and the general public (Calsamiglia & Ferrero, 2003; Calsamiglia & Van Dijk, 2004; Dahlstrom, 2014). Intermediaries such as journalists have played the role of spokespeople for scientists (Weigold, 2001). These in-betweens have been trained to sense the demands of the general public and respond by drawing on their basic understanding of science-related issues. Advances in digital technologies give rise to a second method of scientific popularization. In this method, scientists communicate directly with the lay audience, an approach that is facilitated by an increasing number of digital platforms where scientists can share their knowledge with wide audiences. On these platforms, scientist popularizers can make use of multiple semiotic resources such as images, moving images, audio and video to illustrate their points. Compared to traditional print media, the multimodal resources afforded by digital media can provide visualizations of abstract scientific concepts that enhance scientists' ability to explicate scientific concepts in less technical language. Such multimodal communication practices between scientists and lay people are exemplified by genres like, among others, scientific blogs (Luzón, 2013), science-focused crowdfunding proposals (Mehlenbacher, 2017), online science notebooks (Carter-Thomas & Rowley-Jolivet, 2017), and TED talks (Caliendo, 2014; Scotto di Carlo, 2014, 2015a, 2015b, 2018).

In the process of scientific popularization, readers/audiences are not supposed to be passive recipients of knowledge, but participants actively involved in the process of popularization (Myers, 2003). In view of this, popularizers tend to employ certain discursive engagement strategies, which are techniques used by them to "acknowledge and connect to others" (Hyland, 2010: 125). This definition implies two aspects of audience engagement. Primarily, authors are supposed to be conscious of the existence of the audience or readers; furthermore, they are expected to bond with the audience in some way. This bonding practice includes but may not be limited to "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" (Hyland, 2010: 125). Researchers have found that authors of scientific popularizations employ a variety of strategies to engage the readers, such as the considerable use of inclusive pronouns (Caliendo, 2014; Hyland, 2010; Luzón, 2013; Scotto di Carlo, 2014, 2018), subjective and emotive adjectives (Scotto di Carlo, 2015b), pathos techniques (Scotto di Carlo, 2015a), and questions (Hyland, 2010). In addition to the lexical and syntactic strategies, authors also adopt some strategies at discourse level, for example, a conversational tone (Mattiello, 2017), referencing popular beliefs (Luzón, 2013), and intentional use of some genre moves to solicit social engagement (Mehlenbacher, 2017).

Although these studies provide useful insights into the understanding of engagement in scientific popularization, they tend to focus on written genres. For this reason, they have examined primarily the linguistic mode and rarely gone beyond this to take multimodal resources into consideration. There is therefore a need to better understand the way that combinations of multiple semiotic resources are used to construct meaning in digitally mediated popularization genres. As argued by Luzón (2019: 189), "the strategies to engage the viewers and attain affective engagement are also realized by orchestrating resources of speech (e.g. lexical and grammatical choices which convey intimacy, informality or affinity), image and gestures". Furthermore, research into the use of multimodal resources would help in facilitating scientific popularization online. This is due to the fact that as scientific popularization is increasingly reliant on digital technologies, the attraction of the online audience depends on both linguistic resources and other non-linguistic semiotic resources.

The present study aims to fill the gaps identified above by investigating how multimodal semiotic resources are combined in digital media to engage the online audience in an influential online scientific popularization genre, that is, TED talks, with reference to the following questions: (1) What multimodal semiotic resources are employed in TED talk videos to engage the online audience? (2) What is the relation among these semiotic resources?

2. TED talk videos: digitally mediated scientific popularization practice

One of the most influential online popularization genres is TED talks. In order to disseminate scientific knowledge, TED, a non-profitable

organization, invites scientists, specialists and academics from diverse disciplines to deliver speeches to audiences with diverse backgrounds. Caliendo (2014: 113) argues that TED talks are located "at the intersection between university lectures, scientific communication, newspaper articles, conference presentations and TV science programs". With their skillful combination of various generic resources, TED talks effectively break the traditional triangulating relation between scientists, journalists and the general public, thereby enabling scientists to communicate directly with a lay audience. The genre of TED talks is therefore regarded as "a new and, from the over half a billion views in YouTube and on the TED site, a highly successful form of popularizing discourse" (Sugimoto et al., 2013: 673).

TED talks are characterized by their use of rhetorical strategies to engage the audience. Scotto di Carlo (2014, 2015) and Caliendo (2014) draw upon Hyland's concept of proximity (2010) to analyze speakers' use of linguistic techniques including organization, argument structure, credibility, stance and reader engagement based on their self-constructed corpus of TED talks. Interested in the meta-discourse employed in the texts, Scotto di Carlo (2018) looked into how speakers use different self-mention forms to engage and interact with their audience. Guided by rhetorical appeal theory, in a case study, Scotto di Carlo (2015a) explores speakers' use of *pathos* as a communicative strategy to persuade their audience by arousing their emotion and passion.

These studies provide insights into how TED talks, as a new form of popularization, disseminate knowledge effectively by aligning with the audience. However, it is noteworthy that the audience of these talks includes not only people in the lecture hall, but also millions of netizens worldwide who are able to access the videotaped speeches on the Internet. These online videos make use of not only linguistic resources (i.e., the speech delivered by the speaker), but also digital-multimodal resources such as the image on the screen, the camera position and the distance of shot. Such digital-multimodal elements have not been sufficiently taken into account in previous studies. In view of this gap, the present study aims to investigate how digital-multimodal resources are employed in TED talk videos to engage the online audience.

3. Multimodality

Multimodality is an approach that "understands communication and representation as more than language and attends systematically to the social interpretation of a range of forms of making meaning" (Jewitt, 2016: 69). Regarding communication as an event involving the use of multiple semiotic resources, multimodal analysis foregrounds the investigation of the combination, interplay, or "semantic trade" (Kress, 2000: 339) among various modes in the meaning-making process.

A primary assumption in multimodality is that modes have different affordances. As argued by Kress (2010: 79), "different modes have different potentials for making meaning. These differing potentials have a fundamental effect on the choice(s) of mode in specific instances of communication". For instance, speech and image have different affordances and thus means of making meaning. The affordances of speech include vocal intensity, pitch and vocalic quality. Therefore, the conveyance of meaning through this mode is realized by one sound after another articulated in a sequence of time. The mode of image, on the other hand, includes visual composition, spatial organization, color, and other visual effects; for this reason, the construction of meaning depends on the spatial arrangement of elements presented simultaneously on a surface such as a page or a screen. Due to the different affordances, the two modes have different ways of shaping meaning. For example, to convey the meaning of "sadness", the mode of speech may employ the resources of slow rhythm and low pitch while the mode of image may employ visual resources such as dark color to depict a gloomy scene.

As modes have different affordances, representations and interactions in real communicative settings tend to involve more than one mode. Multimodality scholars regard these communicative events as "multimodal ensembles" (Jewitt, 2016: 72). In a multimodal ensemble, meaning is made through the combination of all available semiotic resources, instead of through one privileged semiotic mode. However, this does not mean that the contribution made by each mode is even. In some scenarios, one mode could play a more dominant role in making meaning than others. Furthermore, as each mode has its specialized functions, one resource is only able to convey part of the meaning and a complete story is more likely to be told through the combination of various semiotic resources.

The concept of multimodal ensemble implies that multimodal research needs to not only investigate the affordances of individual modes, but also

investigate the relation among them. Regarding the interrelation, an idealistic assumption is that all semiotic resources used in a multimodal ensemble would be arranged in a way that harmoniously serves a unified communicative purpose. In reality, though, the functions served by different semiotic resources are not always convergent. Multimodal researchers have identified different types of relation among modes, with Unsworth (2008) describing three types: c*oncurrence,* in which the meaning of each mode reinforces each other; *complementarity*, in which the meanings conveyed by modes are different but supportive of each other; and *divergence,* in which meaning in one mode contradicts meaning in another. In addition, the systemic functional linguistic approach to multimodality (Jewitt, Bezemer & O'Halloran, 2016) identifies two interrelations between semiotic resources: *inter-semiosis* describes the scenario in which semiotic choices interact with each other and are combined; and *re-semiotisation* depicts the situation in which semiotic choices are reconstructed and re-contextualized across modes.

The present study is informed by the theoretical framework of multimodality described above. In the investigation of the engagement strategies used in TED talk videos, attention is paid to the full range of semiotic resources used in engaging the online audience and how these resources interact with each other to serve this purpose.

4. Data sources and analytical methods

The present study aims to investigate how multimodal resources are employed in TED talks, a digitally mediated popularization genre that aims to engage the online audience. This study draws on data from a larger research project that has sampled twenty-eight highly viewed TED talks in the field of biology from 2003 to 2018 (two per year). For the purposes of the present study, multimodal discourse analysis was performed on the talk delivered by Dr. Jill Bolte Taylor at the 2008 TED Conference. Focusing on a single case in this way allows for an in-depth analysis of the semiotic resources deployed as the scientist describes functions of the human brain by recounting her experience of a major stroke. As it deals with the subject of biology, a highly technical discipline, studying this talk can assist in the understanding of how the speaker engages with the online audience in the popularization of a type of science that is regarded as distant from non-specialists. Moreover, this

talk, one of the Top 20 most viewed videos on the TED official website, has a high level of popularity, statistically evidenced by the large number of viewers, approximately 23 million at the date of investigation. Since the main aim of this study is to explore how multimodal resources are combined to engage the audience, taking a best practice approach, this study explores how multimodal resources are combined to engage the audience in this highly successful exemplar.

Once the video was identified, it was firstly viewed several times and initial observations were made. As stated by Jewitt (2016: 77), "multimodal analysis involves an intensive engagement with data. Viewing the data through the lens of research questions serves to refine or generate new questions, clarify criteria for sampling". For this reason, the targeted video was then viewed repeatedly until a holistic sense of the modes and semiotic resources used was gained. After this, thorough intense viewing, multimodal transcription of the whole video was undertaken, following similar analytical procedures to those reported in Baldry and Thibault (2006) and Hafner (2014, 2015). The video was firstly divided according to shots, with one shot defined as a "continuously filmed stretch of video, without cuts" (Hafner, 2014: 666). (Author a). For each shot, the following dimensions were transcribed:

- speech
- represented participant(s) (i.e., the speaker, the visual, the audience, or a combination)
- use of visuals (if applicable)
- size of frames (i.e., distance of shots, including close shot, medium shot, long shot and very long shot)
- height of shots (i.e., low, eye-level, high)
- horizontal angles of shots (i.e. frontal, oblique)
- gaze performed by the represented participant
- gestures performed by the represented participant.

The multimodal transcription procedure was facilitated by Multimodal Analysis Video (Multimodal analysis company Pte Ltd, 2018), a software program that allows users to transcribe the use of multiple semiotic resources into strips and to visualize the proportion of "semiotic interactions", that is, the dynamic interrelation between semiotic resources (O'Halloran, Tan & E, 2015: 17). Figure 1 illustrates the analysis window of the software.

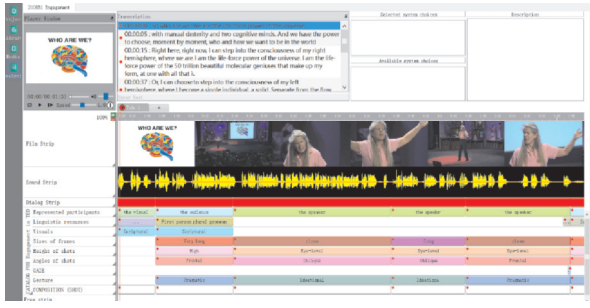

Figure 1. Screenshot of analysis window in Multimodal Analysis Video.

For the purposes of this article, the findings of a segment (16'44"-18'17") that was found to serve the function of engaging the audience are reported. The selection of this segment emerged through the exhaustive scrutiny of the talk and the multimodal transcription procedure.

5. Multimodal engagement in scientific popularization

The selected extract analyzed in this article contains seven shots altogether and lasts for 86 seconds. The multimodal transcription of this extract reveals ten different configurations of semiotic resources as illustrated in Figure 2, which details each configuration including duration of time elapsed. Out of the ten configurations identified, five critical moments, i.e. moments in which semiotic resources are marshalled to special effect, can be identified. We examine these in more detail below.

These ten configurations are of different lengths with some longer than others.

Types of configuration	Duration (seconds)	Represented participant (s) and the combination of resources	Percentage of time (%)
A	4	Visuals, first person plural pronoun, rhetorical questions	5.3
B	10	Speaker + visuals + audience, first person plural pronoun, very long shot, high, frontal angle, gesture	11.6
C	20	Speaker, close shot, eye-level, oblique angle, gesture	22.1
D	9	Speaker, long shot, oblique and eye-level shot, gesture	10.5
E	12	Speaker, close shot, eye-level frontal shot, gesture	13.7
F	1	Speaker, second person pronoun, rhetorical questions, close shot, eye-level shot, frontal angle, gesture	1.1
G	1	Speaker, second person pronoun, rhetorical questions, close shot, eye-level, frontal, gesture, direct gaze	1.1
H	9	Speaker + visuals + the audience, second person pronoun, rhetorical questions, very long shot, high and frontal angle, gesture	10.5
I	16	Speaker, first person plural pronoun, medium shot, eye-level and oblique angle, gesture	17.9
J	4	Speaker, first person plural pronoun, medium shot, eye-level and frontal angle, gesture	5.3

Figure 2. Multimodal configurations in the selected extract.

5.1. Multimodal configuration one: Engagement through visual aids

This configuration takes place in the first shot of the analyzed segment (configuration A), accounting for 5.3% or 6 seconds in terms of the whole film. In this configuration, audience engagement is achieved by verbally asked questions that involve the use of inclusive *we*, a question presented in text echoing the speech, and a graph highlighting the key information conveyed by the speech. As shown in Figure 3, in this shot, one slide is presented in full screen. This slide consists of textual and figurative elements. The heading, namely, the question "WHO ARE WE?" is presented in block letters in bold font. The image presents an abstract representation of a human brain. This image of the human brain is aesthetically and artistically reproduced, depicting a human brain filled with multiple bright and appealing colors. This slide is accompanied by the speaker's speech *"So who are we? We are the life-force power of the universe"*. In this utterance, the speaker asks a rhetorical question ("*so who are we*?"). By asking this question, the speaker invites the audience to contemplate the nature of humanity. This question is immediately followed by the answer ("*We are the life-force power of the universe*"). By using the inclusive *we,* the speaker manages to define the properties ("*life-force power of the universe*") as a commonality between the speaker and the audience, thereby aligning herself with the audience.

It can be seen that semiotic interaction takes place between the visual and the speech. The writing on the slide presents the question; at the same time, the

speaker verbalizes the question again in the speech. This echoing appears to strengthen the invitation sent to the audience and motivates them to consider the question. Furthermore, from the perspective of the online audience, the speaker plays the role of a backstage narrator because the moment is not captured in a way that both speaker and the slide are simultaneously presented. Instead, the online audience is visually exposed to the scripted question and acoustically attracted by the question uttered by the speaker, which could attract the attention of the online audience.

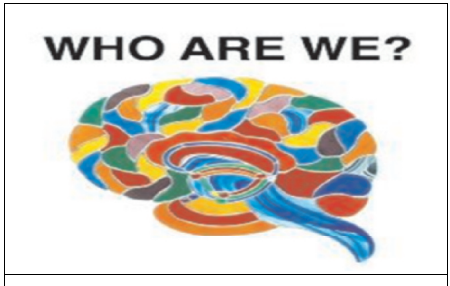

Figure 3. Extract from TED talk "My stroke of insight", 16 min 44 sec -16min 48 sec. Image used with permission.

5.2. Multimodal configuration two: Engagement through a long shot

The second multimodal configuration identified to engage the online audience takes up 11.6% or 10 seconds in terms of the whole extract (configuration B). Different from *Configuration One* when the slide is the only represented participant, at this moment, the represented participants include the speaker, the visual and the audience appearing simultaneously. In order to capture all three participants, a very long shot from a frontal, high angle was adopted (see Figure 4). The use of long shot, as suggested in the "Camera Playbook: Shooting practice" (TED, 2012), can capture the speaker, the projection screen and sometimes the audience so that the online viewers get a sense of what the live experience is like. This is rather counter-intuitive as an involvement strategy: while one would normally associate a long shot with far social distance, what it actually does in this instance is to include online viewers in the auditorium, giving them the sense that they are a part of the audience and therefore involved. A high angle can also facilitate audience engagement, compensating for the absence of online viewers from

the conference venue. Such a high angle enables the online viewer to tower high above the conference venue thereby exercising a symbolic power over the participants represented in the video (Kress & van Leeuwen, 2006). Furthermore, the use of frontal angle further engages the audience as frontal angle, with gaze fixed on the camera, in a demand shot. According to Kress and van Leeuwen (2006: 122), "'demand' images address the viewer directly, realizing a visual 'you"'. Therefore, through the employment of the frontal shot, online viewers can gain a sense of being directly addressed by the speaker, even though far away.

The speaker also contributes to the engagement of the online viewers by deploying additional semiotic resources (i.e. verbal language and gesture) at this moment. Verbally, she employs first person plural pronoun to refer to herself, the audience, and humanity in general, thereby creating an intimate relation between herself and the audience. Besides, the utterance "*we have the power to choose, moment by moment, who and how we want to be in the world*" is accompanied by a recurrent gesture. The speaker had her loose hand moving towards her body back and forth. This is a gesture that serves the illocutionary force of showing assertiveness (Bressem & Müller, 2014) when the speaker was stating a feature possessed by both the speaker and the online audience.

Shot 2 (16min 49 sec)

Shot 2 (16min 55sec)

Shot 2 (16min 59 sec)

*with manual dexterity and two cognitive minds. And **we** have the power to choose, moment by moment, who and how **we** want to be in the world.*

Figure 4. Extract from TED talk "My stroke of insight", 16min 49 sec – 16 min 59 sec. Images used with permission.

5.3. Multimodal configuration three: Engagement through gaze

Another multimodal engagement took place at 17 min 42 seconds (configuration G) of the talk (see Figure 5). Although this configuration takes up only 1.6%, or 1 second in terms of the whole film, a wide range of

semiotic resources, that is, the frontal, eye level, medium shot, gesture, the second person pronoun, direct question and gaze, are used in combination to engage the online audience multimodally. The represented participant, that is, the speaker, is captured with a medium shot from an eye-level, frontal angle. An eye-level shot creates a point of view of equality and there is no power difference between the speaker and the viewer. A frontal angle is also used to convey a sense of involvement. At the same time, an above-waist medium close shot tends to establish a comfortable distance between the represented participant and the viewers, somewhere between a polite social distance and personal intimacy (Kress & van Leeuwen, 2006). The language used at this moment is also engaging. In the short utterance "*which would you choose?*", two engagement techniques are employed. Firstly, the speaker addresses the audience directly by using the second person pronoun *you*. A direct address like this allows the speaker to have a direct conversation with the audience. More importantly, this direct address is embedded in a direct question asking the audience to consider their own choice. While uttering this question, the speaker used a gesture resembling a welcome gesture, seeming like an invitation to reflect on the question raised.

These semiotic resources are further complemented by the use of a direct gaze. In the case of the TED Conference, TED speakers spend most of the time maintaining eye contact with the audience in the conference venue. They may not pay particular attention to the online audience, hidden as they are behind cameras. However, in Figure 5, the speaker appears to gaze directly into the camera, conveying an impression of involvement through gaze directed at the online audience. According to Kress and van Leeuwen (2006), a direct gaze conveys a demand that invites the viewer to enter into some relation with the represented participant. Therefore, by being directly gazed at, the online audience may gain the impression that they are powerfully involved as part of the scene depicted.

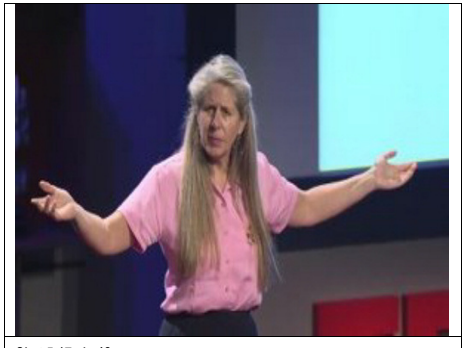

Figure 5. Extract from TED talk "My stroke of insight", 17min 42 sec.
Image used with permission.

5.4. Multimodal configuration four: Engagement with questions

Configuration H contains a critical moment of multimodal engagement, as illustrated in Figure 6. In this shot, the speaker asked two questions following the preceding question in Shot 5 ("*Which would you choose*?"). By asking these direct questions, the speaker invites the audience to consider the issues raised and therefore motivates them to participate in the process of contemplation. These questions are combined with the use of second person pronouns ("*Which do you choose*?"). In this way, the audience is positioned as active participants who are invited to think over the questions that are closely related to human destiny. Another feature of the questions is that they appear in parallel repetition ("*Which would you choose? Which do you choose? And when?"*). This use of parallel questions reinforces the engaging request for information and urges the audience to consider the raised issues. When these questions were uttered, the speaker intentionally paused between the two questions, which can be seen from the sound strip screenshot (Figure 7). The pauses here can engage the audience by allowing some time for them to think (Grice, Skinner & Mansson, 2016). The visual resources employed in this moment are similar to *Multimodal Configuration Two* when a long shot with high and frontal angle is adopted. By taking this shooting angle, an online viewer is empowered and given a sense of the live experience. At the same time, by using the long shot, the online audience is allowed to view the slide that also serves the function of engaging the audience verbally and visually. This combination of a wide range of semiotic resources can effectively involve and engage online viewers.

Shot 6 (17min 43 sec)	Shot 6 (17min 48 sec)	Shot 6 (17min 52 sec)
Which do you choose? And when?		

Figure 6. Extract from TED talk "My stroke of insight", 17min 43 sec -17min 52 sec. Images used with permission.

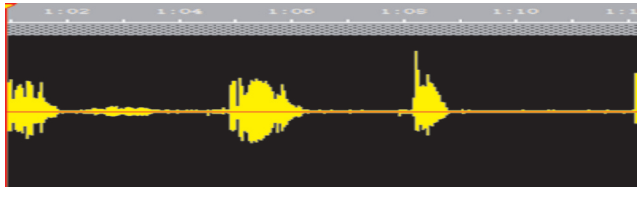

Figure 7. Sound strip of the video segment between 17min 42 sec and 17min 52sec, shots 5 and 6.

5.5. Multimodal configuration five: Engagement through personal emotions

The last multimodal configuration analyzed takes place from 17min 57sec at the video, lasting for 16 seconds. This configuration state accounts for 17.9% in terms of the whole extract. During this 16-second segment, the speaker expressed her expectations of herself, the audience and humanity and motivated the audience to search for inner serenity based on the renewed understanding of the anatomy of the human brain. In so doing, the speaker went beyond the neutral, impersonal introduction of scientific knowledge to the expression of personal emotions. In this process, the speaker presents herself as a sympathetic peer rather than a scientist with a neutral, non-emotional stance. Verbally, the speaker employs the first person plural pronoun *we* to align with the audience in terms of sparing efforts in finding inner peace ("*the more time **we** spend choosing to run the deep inner-peace circuitry*") and achieving a satisfying outcome ("*the more peace **we** will project into the world*"). The inclusive possessive pronoun *our* is also used by the speaker to bond with the audience. By qualifying the "right hemisphere" and "planet" as belonging to both the speaker and the audience, the speaker manages to strengthen the sense of community. Accompanying the linguistic techniques,

semiotic resources of gesture are also adopted. When the speaker utters "*the more time we spend choosing*", she pulls her hands (starting arms outstretched at her sides) towards her upper body with fingers pointing to herself (see the screenshots at 17min 53 sec and 17min 59 sec). This gesture appears to strengthen the sense of "we". This inward pointing gesture lasts until she utters "*will project into the world*" with her hands pushing towards the air in front of her upper body. In addition to these resources contributed by the speaker, a medium close shot is adopted by the production crew to create a comfortable distance between the speaker and the audience without being too personal or too distant. At the same time, an eye level shot is adopted to convey an equal power relation between the online viewers and the speaker.

Figure 8. Extract from TED talk "My stroke of insight", 17min 53 sec - 18min 09 sec. Images used with permission.

6. Discussion

While the findings of this study are drawn from a limited amount of data, it is nevertheless possible to generate some insights into online scientific popularization practice. The analysis shows that online scientific

popularization tends to go beyond the introduction of scientific knowledge to the construction of audience engagement. In the present case study, it is found that the scientist speaker did not only disseminate scientific knowledge in the speech, but also devoted time to relating the scientific issue to the audience by employing various engagement strategies. A possible reason for doing this is summarized by Anderson (2016: 47) in the official guide to giving a TED talk who states that "you can give the most brilliant talk, with crystal-clear explanations and laser-sharp logic, but if you don't first connect with the audience, it just won't land". This statement indicates that the necessity of creating the connection with the audience results from the need to convey the speakers' ideas to the audience. This means that audience engagement is not the end, but a means to the end, that is, effective popularization of science among the audience.

This understanding of the importance of audience engagement challenges the dominant view of scientific popularization which places specialists and lay people at two extreme ends of the popularization process and regards scientists as the dominant player in the practice (Gregory & Miller, 1998; Myers, 2003). Instead, the engagement strategies adopted by the scientist speaker indicates that in online popularization practice the audience is expected to be part of the popularization. For example, when the speaker employed first person plural pronoun (i.e., the inclusive *we*), the possessive adjective *our* and the question directly addressing the audience, the audience is likely to feel involved and perhaps even motivated, rather than passively accepting an idea implanted in their mind. In order to with the audience, the scientist needs to take on the role of not only a specialist but also a guide with the willingness to interact with the audience. As argued by Gotti (2014: 23), "popularization is thus not just seen as a category of texts, but as a recontextualization process that implies relevant changes in the roles taken on by the actors and institutions involved, and their degree of authoritativeness".

The findings demonstrate that engagement in a digitally mediated popularization genre like TED talks is achieved not by one single semiotic mode, but by the combination of various modes, such as visuals, text on screen, distance of shot, perspective of shot, shooting angles, gaze, gestures and speech. The analysis reveals five different multimodal configurations of these semiotic resources to achieve viewer engagement, namely, engagement through visual aids, the long shot, the gaze, questions and personal emotions. Note that, while we have focused our analysis on particular forms of

expression at particular moments, this is not to downplay the important contributions made through other modes simultaneously. It is also worth noting that the strategies can be overlapping because a specific moment could involve more than one strategy.

The findings also demonstrate that the interrelations among the semiotic resources employed to engage the audience can vary. Firstly, the findings indicate that multiple semiotic resources can be combined to serve the same communicative purpose of engaging the online viewers. This type of relation is described by Unsworth (2008) as "concurrence", indicating that the meaning of each mode reinforces the meaning of others. An example of this is *Configuration One: Engagement with Visual Aids*. In this configuration, when in the speech a question is raised verbally to address the audience directly, the slide at the same time demonstrates the question in special font and size to strengthen the communicative effect of attracting the audience's attention. Here, the speech and writing reinforce each other. This pattern of multimodal orchestration can also be seen in *Configuration Three: Engagement through Gaze*. In this configuration, a medium close shot is adopted to establish a close social distance between the viewer and the speaker; a frontal angle and an eye-level shot are used to create an equal and involving relation between the two parties; at the same time, the welcoming gesture, the direct gaze and the linguistic resources all function to engage the audience. In both cases, interpersonal meanings are expressed through a range of modes concurrently.

The second type of relation identified in the data is when one semiotic resource is unavailable at a specific moment, other resources can compensate and maintain the engaging effect. This can be exemplified by *Configuration Five: Engagement through Personal Emotion*. In this configuration, the speaker is depicted from an oblique angle. The use of oblique angle tends to generate the interactive meaning of detachment instead of engagement (Kress & van Leeuwen, 2006). This way of using a non-engaging shooting angle suggests that at this moment not all multimodal representation choices necessarily reinforce one another in a uniform way. However, although the horizontal angle of shot at this moment does not engage the audience, other semiotic resources nevertheless serve the function of engaging them, for instance, the above waist upper body shot corresponding to near social distance, the eye level shooting height and the linguistic techniques. This type of relation is closer to what Unsworth defines as "complementarity" (2008) among semiotic resources. In this way, the video is able to present subtly different

social distances in different configurations, while preserving an overall stance that is engaging and involving.

This case study also reveals the feature of multiple authorship involved in the construction of this digitally mediated popularization genre. In the present case, the scientist speaker contributes a range of semiotic resources including the embodied speech, the slides, use of gesture and gaze to engage the online viewers. Furthermore, as the speech is captured in the format of a video, visual resources such as the size of frame, the horizontal perspectives and the height of shots can also convey interpersonal meaning. Selection of these visual resources tends to be made by the production team. This feature of collaborative authorship is also found in other digital genres such as video methods articles (Hafner, 2018), a(Author c) and teamwork scientific documentaries constructed by undergraduate students (Hafner, 2014). (Author a). The involvement of multiple authorship may be due to the fact that the meaning making practice in such complex digital-multimodal ensembles depends on multiple semiotic resources. Assembling such a wide range of resources likely requires diverse skill sets and competencies. As a result, the semiotic resources employed to achieve particular communicative purposes (e.g., engaging the online audience) are likely to be contributed by multiple professionals with different expertise.

Regarding the relation among these authors, the case study shows that the semiotic resources contributed by one author may influence the semiotic choices made by another author. An instance of this is *Multimodal Configuration Three: Engagement through Gaze* when the speaker employs engaging gestures and direct gaze. Such semiotic choices tend to encourage the production crew to choose medium close, eye-level and frontal shots to reinforce the interpersonal meanings in the speaker's talk. In addition, if a long shot or an oblique shot was employed at this moment, the performance of the speaker would not be fully captured and consequently the meaning conveyed by the gesture and the gaze may not be effectively conveyed to the online viewers.

The findings also support the process of *resemiotization* in digital-multimodal texts. This concept is raised to depict the process by which spoken and interactional text gets encoded, re-semiotized and ultimately solidified into more permanent structures (Iedema, 2001). This process can be reflected in the TED talk video. At the beginning, the scientist contributes semiotic resources including the speech, the slides, the gestures and the gaze. At this

stage, the presentation of the speaker is a temporal meaning-making activity which can only be received once by the limited number of people in the conference venue. After the production crew add digital-multimodal resources such as the distance of shot, the height and the angle of shots, the speech is materialized in the form of a video, which is more durable and stable. The meaning conveyed in this durable, technologized form can be received repeatedly by an unlimited number of online viewers from around the world. This process of resemiotization made possible by the involvement of multiple authors and digital technologies is important for scientific popularization because it solidifies the information conveyed in the speech thereby allowing the message delivered by the scientist to be recycled online permanently.

7. Conclusion and implications

The study demonstrates that the communicative aim of engaging online viewers in a digitally mediated scientific popularization genre like TED talks tends to be achieved by the combination of multiple semiotic resources afforded by digital technologies. In the construction of the text, authors are provided with an expanded range of semiotic choices such as the use of speech, gesture, gaze, visuals, distance of shot, perspective of shot and angle of shot. These digital-multimodal semiotic resources supplement each other or are orchestrated in order to achieve the engagement of the online audience.

All of the foregoing has implications for English for Specific Purposes (ESP) researchers. Many studies of genres continue to adopt a linguistically focused genre analysis, including studies of a digital-multimodal genre like TED talks. In failing to consider multimodal resources, like gesture, editing or visuals, such studies may provide an impoverished analysis. As Carter-Thomas and Rowley-Jolivet (2003: 9) observe, "complete generic perspective on a discourse involves not only considering language, but also taking into account all the semiotic resources brought into play in the given discourse situation". For this reason, there is a need for more of the kind of studies that go beyond the consideration of linguistic resources to take into account the use of digital and multimodal semiotic resources. This article demonstrates a methodological approach that can be applied elsewhere that would allow researchers to account for these different resources. The

software-based approach applied in this study can assist genre researchers to conduct a systematic transcription of the multiple semiotic resources used in digital-multimodal genres and facilitate the identification of multimodal configurations in these genres.

By highlighting the multimodal and digitally-mediated nature of scientific communication in the case of TED talks, this study also points towards an emerging need for science students and novice scientists to develop the necessary skills to communicate with various, diverse audiences through multimodal expression. This need is echoed by recent trends in the academy, where there has been a demand for academics to engage with lay audiences as well as specialists. Therefore, there is a need for ESP pedagogy to foster such communicative competencies. The present study thus supports recent attempts to address multimodality in ESP courses by engaging in educational practices of digital multimodal composition (see Belcher, 2017; Hafner & Miller, 2019, Hafner & Ho, 2020). By involving digital-multimodal elements in language teaching and learning, students can be better prepared for real life communication in this digital era.

In addition, the study shows how collaborative authorship plays an important role in the construction of the TED talk videos. While ESP research has tended to downplay such collaborative processes, the contextual factors that may influence digitally mediated genres have begun to attract the attention of some researchers (Hafner & Yu, 2020; Hynninen, 2018; McGrath, 2016). These studies have pointed to a trend among genre scholars to look at not only the product, that is, the text, but also the process of genre construction. As argued by Kress and van Leeuwen (2001: 49), "in many contexts, reporting, writing, (sub)editing, layout, publishing, are all merging into a single new practice, through the availabilities and affordances of electronic technologies". This new practice of constructing digitally mediated genres calls for further investigation into the design process. Over its nearly fifty-year history, ESP has gradually moved from a focus on the language of the text to an increasing interest in more contextualized factors. This trend allows us to see in detail the kind of activities that are taking place in genre construction, activities that are supported by, and made more complex by, the affordances of digital media. In future, ESP studies should continue to engage with such online genres, examining not only their discursive products but also the collaborative practices that are observed in genre production.

Article history:
Received 21 May 2020
Received in revised form 16 January 2021
Accepted 1 March 2021

References

Anderson, C. (2016). *TED Talks: The Official TED Guide to Public Speaking*. Headline Publishing Group.

Baldry, A. & J. P. Thibault (2006). *Multimodal Transcription and Text Analysis: a Multimedia Toolkit and Coursebook*. Equinox.

Bondi, M., S. Cacchiani & D. Mazzi (2015). *Discourse in and through the Media: Recontextualizing and Reconceptualizing Expert Discourse*. Cambridge Scholars Publishing.

Bressem, J. & C. Müller (2014). "A repertoire of German recurrent gestures with pragmatic functions" in C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill & S. Tessendorf (eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*, 1575-1591. De Gruyter Mouton.

Caliendo, G. (2014). "The popularization of science in web-based genres" in G. Bongo & G. Caliendo (eds.), *The Language of Popularization*, 111-140. Peter Lang.

Calsamiglia, H. (2003). "Popularization discourse". *Discourse Studies* 5(2): 139-146. DOI: <https://doi.org/10.1177/1461445603005002307>.

Calsamiglia, H. & C. L. Ferrero (2003). "Role and position of scientific voices: reported speech in the media". *Discourse Studies* 5(2): 147-173. DOI: <https://doi.org/10.1177/1461445603005002308>.

Calsamiglia, H. & T. A. Van Dijk (2004). "Popularization discourse and knowledge about the genome". *Discourse & Society* 15(4): 369-389. DOI: <https://doi.org/10.1177/0957926504043705>.

Carter-Thomas, S. & E. Rowley-Jolivet (2003). "Analysing the scientific conference presentation (CP): A methodological overview of a multimodal genre". *ASp: La revue du GERAS* 39-40: 59-72.

Dahlstrom, M. F. (2014). "Using narratives and storytelling to communicate science with nonexpert audiences" in *Proceedings of the National Academy of Sciences of the United States of America* 111: 13614-13620. DOI: <https://doi.org/10.1073/pnas.1320645111>.

Gotti, M. (2014). "Reformulation and recontextualization in popularization discourse". *Ibérica, Journal of the European Association of Languages for Specific Purposes* 27: 15-34.

Gregory, J. & S. Miller (1998). *Science in Public: Communication, Culture and Credibility*. Plenum Trade.

Grice, G. L., J. F. Skinner & D. H. Mansson (2016). *Mastering Public Speaking* (9th ed.). Pearson.

Hafner, C. A. (2014). "Embedding digital literacies in English language teaching: Students' digital video projects as multimodal ensembles". *TESOL Quarterly* 48(4): 655-685. DOI: <https://doi.org/10.1002/tesq.138>.

Hafner, C. A. (2015). "Remix culture and English language teaching: The expression of learner voice in digital multimodal compositions". *TESOL Quarterly, 49*(3), 486-509. DOI: <https://doi.org/10.1002/tesq.238>.

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: 15-42.

Hafner, C. A. & W. Y. J. Ho (2020). "Assessing digital multimodal composing in second language writing: Towards a process-based model". *Journal of Second Language Writing* 47: 100710. DOI: <https://doi.org/10.1016/j.jslw.2020.100710>.

Hafner, C. A. & L. Miller (2019). *English in the Disciplines: A Multidimensional Model for ESP Course Design*. Routledge.

Hafner, C. A. & C. Yu (2020). "Language socialization in digitally mediated collaborative writing: Evidence from disciplinary peer and teacher feedback". *RELC Journal* 51(1): 14-32. DOI: <https://doi.org/10.1177/0033688220901347>.

Hyland, K. (2010). "Constructing proximity:

Relating to readers in popular and professional science". *Journal of English for Academic Purposes* 9(2): 116-127. DOI: <https://doi.org/10.1016/j.jeap.2010.02.003>.

Hynninen, N. (2018). "Impact of digital tools on the research writing process: A case study of collaborative writing in computer science". *Discourse, Context & Media* 24: 16-23. DOI: <https://doi.org/10.1016/j.dcm.2018.01.005>.

Iedema, R. (2001). "Resemiotization". *Semiotica* 137(1/4): 23-39.

Jewitt, C. (2016). "Multimodal analysis" in A. Georgakopoulou & T. Spilioti (eds.), *The Routledge Handbook of Language and Digital Communication*, 69-84. Routledge.

Jewitt, C., J. Bezemer & K. L. O'Halloran (2016). *Introducing Multimodality*. Routledge.

Kress, G. (2000). "Multimodality: Challenges to thinking about language". *TESOL Quarterly* 34(2): 337-340. DOI: <https://doi.org/10.2307/3587959>.

Kress, G. (2010). *Multimodality: A Social Semiotic Approach to Contemporary Communication*. Routledge.

Kress, G. & T. Van Leeuwen (2001). *Multimodal Discourse: The Modes and Media of Contemporary Communication*. Arnold.

Kress, G. & T. van Leeuwen (2006). *Reading Images: The Grammar of Visual Design* (2nd ed.). Routledge.

Luzón, M. J. (2013). "Public communication of science in blogs: Recontextualizing scientific discourse for a diversified audience". *Written Communication* 30(4): 428-457. DOI: <https://doi.org/10.1177/0741088313493610>.

Luzón, M. J. (2019). "Bridging the gap between experts and publics: The role of multimodality in disseminating research in online videos". *Ibérica, Journal of the European Association of Languages for Specific Purposes* 37: 167-192.

Mattiello, E. (2017). "The popularisation of science via TED talks". *International Journal of Language Studies* 11: 77-106.

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. DOI: <https://doi.org/10.1016/j.esp.2016.02.003>.

Mehlenbacher, A. R. (2017). "Crowdfunding science: Exigencies and strategies in an emerging genre of science communication". *Technical Communication Quarterly* 26(2): 127-144. DOI: <https://doi.org/10.1080/10572252.2017.1287361>.

Myers, G. (2003). "Discourse studies of scientific popularization: questioning the boundaries". *Discourse Studies* 5(2): 265-279. DOI: <https://doi.org/10.1177/1461445603005002006>.

O'Halloran, K. L., S. Tan & M. K. L. E (2015). "Multimodal analysis for critical thinking". *Learning, Media and Technology* 42(2): 147-170. DOI: <https://doi.org/10.1080/17439884.2016.1101003>.

Scotto di Carlo, G. (2014). "The role of proximity in online popularizations: The case of TED talks". *Discourse Studies* 16: 591-606. DOI: <https://doi.org/10.1177/1461445614538565>.

Scotto di Carlo, G. (2015a). "Pathos as a communicative strategy for online knowledge dissemination: The case of TED talks". *3L, Language Linguistics Literature, The Southeast Asian Journal of English Language Studies* 21(1): 23-34.

Scotto di Carlo, G. (2015b). "Stance in TED talks: Strategic use of subjective adjectives in online popularization". *Ibérica, Journal of the European Association of Languages for Specific Purposes* 29: 201-221.

Scotto di Carlo, G. (2018). "Patterns of clusivity in TED Talks: When 'you' and 'I' become 'we'". *Ibérica, Journal of the European Association of Languages for Specific Purposes* 35: 119-144.

Sugimoto, C. R., M. Thelwall, V. Larivière, A. Tsou, P. Mongeon & B. Macaluso (2013). "Scientists popularizing science: Characteristics and impact of TED talk presenters". *PLOS ONE* 8(4): 1-8. DOI: <https://doi.org/10.1371/journal.pone.0062403>.

TED (2012). *Camera Playbook: Shooting Practice.* In TED (ed.). URL: <https://pb-assets.tedcdn.com/system/baubles/files/000/004/585/original/CameraPlaybook-2012-Small.pdf?1484062157>.

Thomas, G. & J. Durant (1987). "Why should we promote the public understanding of science?". *Scientific Literary Papers* 1: 1-14.

Weigold, M. F. (2001). "Communicating science: A review of the literature". *Science Communication* 23(2): 164-193. DOI: <https://doi.org/10.1177/1075547001023002005>.

Sichen Ada Xia is Lecturer at the School of Humanities and Social Science (HSS), The Chinese University of Hong Kong, Shenzhen. Her research interests include English for Specific Purposes, genre analysis, multimodality, scientific popularization and digital literacy.

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 books include *Understanding Digital Literacies: A Practical Introduction* (2nd edition) (Routledge, 2021, co-authored with Rodney Jones), and *English in the Disciplines: A Multidimensional Model for ESP Course Design* (Routledge, 2019, co-authored with Lindsay Miller).