Maria Elisa Fina
University of Salento, Italy

VOICING CITIES: INTERACTION OF VOICES IN CITY AUDIO GUIDES IN ITALIAN AND IN ENGLISH

Abstract

Although multimodality in tourist communication has been widely investigated, published research so far has focused mainly on the interaction between text and pictures in printed and web-based tourist genres, while little research has been carried out on the aural dimension of audiovisual tourist texts. This study proposes a multimodal analysis of fifty city audio guides in Italian and in English, aimed at exploring how different 'voices' may be combined to create meaning and involve the listener. The analysis focuses on the types of speaker involved in narration and the different ways in which they interact. Furthermore, following van Leeuwen's model (1999), it also shows how different semiotic resources are combined to enhance narration. The results show significant differences between audio guides in Italian and in English, and also bring to light cases of hybridization in the audio guide genre. These issues are taken as starting points for pedagogical implications for students and professionals in the field.

Keywords: city audio guides, complex medium, genre embedding, soundscape

1. Multimodality in tourism studies

The interaction of different semiotic resources has been widely exploited in tourist communication in order to create eye-catching and persuasive texts (Antelmi, Santulli, & Held, 2007), and over the last decade numerous studies have been carried out to explore how multimodality is used to create meaning and produce persuasive effects on the audience. So far, most studies have focused on the interaction between text and pictures in tourist guides (Antelmi et al., 2007; Denti, 2007), printed advertisements (Maci, 2013; Santulli, 2007), websites (Francesconi, 2016; Maci, 2007, 2013; Manca, 2016), brochures (Francesconi, 2011a; Ling Ip, 2008), postcards (Francesconi, 2011b), and blogs (Denti, 2015), to name just a few. Audiovisual tourist texts have been investigated to a far lesser extent. Seminal works in this domain are those by Francesconi, who carried out multimodal analysis of a promotional video of New Zealand (2011c), and of creative aspects in a digital diary video on the Basilicata region (2015, 2017). She also carried out aural analysis by investigating 'soundscapes' in tourist multimodal texts, looking at how speech, music and sounds interact to create meaning and persuade the listener in a radio commercial promoting India, and investigating prosody in a travel radio programme (2014).

As far as audio guides are concerned, the published literature is scant, limited to specific areas of investigation, and almost inexistent from the point of view of multimodality. This lack of research on audio guides has been attributed by Tallon (2006) to criticism over audio guides, whose capability to mediate art interaction is often questioned by art historians. However, it could also be attributed to the much longer tradition of studies on written genres, which are more likely to meet the needs of a wider range of users. Since the early 2000s, substantial research has been devoted to usability features of museum audio guides. Studies in this area basically propose innovative models and systems for increasing the usability of context-aware mobile devices for museums, as in Petrelli and Not (2005), Luyten and Coninx (2004), Gebbensleben, Dittmann, & Vielhauer (2006). Relevant to multimodality is a study by Alfaro, Nardon, Pianesi, Stock, & Zancanaro (2005), who propose solutions to enhance audio presentation in museum multimedia guides in order to allow the audio guide user to personalise and interpret the information according to his/her pace and interests. The solution proposed by the authors involves synchronising a sequence of pictures with the audio commentary and planning the transitions between pictures according to cinematic techniques, which include sequences of camera movements applied to the same image and transitions effects among shots, as opposed to the traditional slide show with no zooming in to display particular details, nor transition effects to create continuity. Relevant to multimodality are also studies by Neves (2012, 2016), in which she addresses the crucial issue of audio description in museums for non-sighted people. She proposes 'soundpainting' as a form of artistic transcreation (2012), in which carefully chosen words and a careful direction of voice with adequate tone and rhythm can work together with music and sounds to build the stories and emotions that a piece of artwork may offer. She also proposes the case of an "enriched descriptive guide" (2016), by which the user's multisensory experiences are stimulated by providing thinking prompts that invite cognitive and/or physical exploration and capture the uniqueness of the cultural context the guide relates to.

The study presented in this paper is part of a wider cross-language, multimodal investigation of the audio guide genre in Italian and in English (Fina, 2016) which involved different levels

of analysis: structure and content types on a macro-structural level; language and style on the micro-structural level; music, sounds and prosody on the multimodal level. These aspects were investigated from the points of view of accessibility and visitor involvement, highlighting similarities and differences between audio guides in Italian and in English and defining best practices for audio guide creation by referring to the framework provided for public speaking, and focusing in particular on the principles of accessibility and listener involvement. The aim of this paper is to evidence how different voices are employed and combined in audio guides to create meaning and involve the listener and, while doing so, bring to light possible differences between Italian and English. The concept of 'voices' here is intended in its broader sense and, in order to be fully understood, a few theoretical premises will be discussed in the following sections.

2. Theoretical background

Before proceeding with the analysis, it is necessary to describe the semiotic nature of the audio guide text in relation to all possible elements that may come into play. The study carried out in this paper draws on two theoretical models that can be used to describe the audio guide as a genre under three different aspects: medium, genre hybridization and interaction of multimodal resources. In the following sub-sections, we will present the key features of these models, and we will relate them to the audio guide. We will first discuss the stylistic nature of the audio guide genre in reference to the written vs. spoken divide; then, we will place the audio guide genre within van Leeuwen's model of 'soundscape' (1999), and, finally, we will define the 'voices' that will be investigated in the audio guide soundscape.

2.1. The hybrid stylistic nature of the audio guide genre

Crystal and Davy (1969) investigated the discourse of medium by looking at the distinctive features of speech and writing in reference to the classification of texts in the Survey of English Usage (SEU) (Quirk, 1959), which is based on the fundamental distinction between speech and writing. The two scholars focused their attention on the "undesirable asymmetry" (Crystal, 1994, p. 36) arising from the fact that not always does language stay in one category (either written or spoken), but there are cases in which switch occurs as, for example, in dictation, where speech is produced to be written down.

The audio guide text is characterised by what Crystal and Davy term complex medium in order to classify "certain features of a variety which would fall as a general rule within one kind of discourse by reference to features which would normally be expected to occur only in another" (Crystal & Davy, 1969: 71). In a previous study on the use of pauses in audio guides (Fina, in press) the audio guide was classified as a complex medium text, conflating specific features that can be seen in Figure 1 (see emphasized elements):
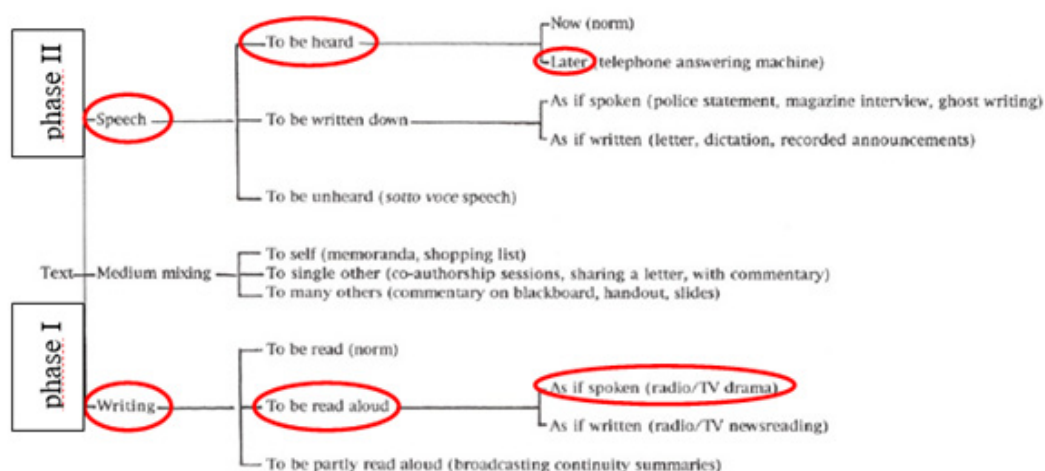
Figure 1. Classification of the audio guide as a complex medium text

In Crystal's words (1994), the production of an audio guide would be the case in which we choose to "write with the intention that what we have written should be read aloud. We may write in such a way that our end-product, when read aloud, will sound like written language. It will be relatively formal and controlled" (p. 39). As a complex medium text, then, the audio guide can be configured as a written text to be read aloud in phase I, while the final product of phase II will be a recorded speech, which will be delivered to final receivers asynchronously, at a later stage, i.e. while sightseeing or in any other moment following publication and downloading of the audio guide.

This hybrid stylistic nature of the audio guide certainly has implications in terms of effective delivery. Given its mainly informative nature, the contents of an audio guide script may be more or less complex depending on the slant given to the guide. The wider investigation of the genre (Fina, 2016) showed that audio guides are dense with historical information and artistic descriptions, with a more or less frequent use of specialized language, mainly from the fields of architecture, geography, geology, and similar. These complex contents, though, are to be delivered orally, and this raises the crucial issue of effective delivery, which can be ensured only by adopting strategies for visitor involvement. Indeed, as founder of *Audyssey Guides*[1] Rob Pyles notes (Holt, 2007), visitors have a very short attention span[2], and considering that audio guides may also last over 120 minutes, the need for adequate strategies for effective delivery becomes particularly significant. In practice, the script is supposed to be made fit for oral delivery by adopting a series of verbal and extra-verbal strategies that make narration more pleasant to be listened to. In this study, these strategies will be investigated.

1   Audissey Guides is an award-winning American producer of audio guides. The website is http://www.audisseyguides.com/about/ (last accessed February 2015).

2   http://matadornetwork.com/notebook/how-to-make-an-audio-tour-ten-tips-from-audissey-guide-pioneer-rob-pyles/ (last accessed February 2016).

2.2. Genre hybridization and 'interdiscursivity'

The 'complex medium' model illustrated in the previous section can be applied to most of the audio guides collected for the investigation, but not to all. As will be shown in the analysis, a small number of audio guides fall outside this model as they also include spontaneous speech delivered by people with special knowledge of the site. Hence, they can be considered as written scripts to be read aloud only in part, since the spontaneous speech is presumably recorded separately and then inserted in the relevant points of the script. There may also be audio guides entirely recorded on site, which cannot be described as written texts to be read aloud at all. These types of audio guide are extremely meaningful for the purpose of this paper, but in order to be taken into account they need to be framed within a different theoretical model. We will relate this particular aspect of the audio guides to genre studies, and more specifically to what Bhatia defines as "interdiscursivity", that is "innovative attempts to create hybrid or relatively novel constructs by appropriating or exploiting established conventions or resources associated with other genres and practices" (2010, p. 35). Interdiscursivity implies a variety of discursive processes and professional practices by which genres are dynamically configured, and often results in "mixing, embedding, and bending of generic norms in professional contexts" (*ibid.*). Genre embedding involves using a genre "as a template to give expression to another conventionally distinct generic form" (1997, p. 191), as for example an advertisement written as if it were a poem, but still recognisable as an advertisement or, in travel texts, a comic strip embedded in a postcard (Francesconi, 2014, p. 29). Genre bending involves "a process of adaptation of an existing genre giving rise to a new genre with different communicative purpose, through a process called 'repurposing'" (*ibid.*), as in the case of Wikitravel, in which the Wikipedia encyclopaedia format has been bended to offer an open digital travel guide (*ibid.*). Finally, in genre mixing the genres involved are no longer distinguishable, as in 'docufiction' (*ibid.*). In this study, we will show how the hybrid nature of the audio guide genre resides not only in its complex medium, but also in the inclusion of features belonging to other genres, and we will illustrate which of the interdiscursivity processes are enacted.

2.3. The 'soundscape'

The theoretical model used to investigate audio guides from a multimodal perspective is van Leeuwen's 'soundscape' (1999), defined as a composite semiotic system made up of speech, music, and sounds. Everyday life offers countless situations in which we may find ourselves immersed in soundscapes, where different types of sound interact at different degrees of loudness. Soundscape is "rarely fixed and static [as] the presence, relevance and interaction of sounds constantly change over time, either spontaneously or as a result of sound manipulation" (Francesconi, 2014, p. 109). Clear examples of sound manipulation are films, where diegetic sounds are often amplified or reduced to mark specific aspects of the scene, while extra-diegetic sounds are added to impact the audience's emotions. By 'sound', however, we do not mean sounds coming from external sources only. Sound also relates to the way

we use voice. In fact, in either formal or informal contexts, we tend to adjust voice pitch, loudness and intonation according to what is required by the situation (e.g. in a library, voice loudness must be low) or depending on the effect we may want to produce on the listener. Therefore, sounds are important because they create meaning and produce effects.

Van Leeuwen's model consists of an analytical framework which aims to investigate speech, music and sounds as interrelated phenomena, in terms of "sound-as-sound", "sound-as-music" or "sound-as-language" (van Leeuwen, 1999, p. 6). Indeed, the model aims to "*integrate* speech, music and other sound" (1999, p. 4, emphasis in the original) and provides specific terminology for describing the integration of the three elements in all its potentials. Van Leeuwen brings as examples a variety of sound events including popular songs, radio and television programmes, commercials and soundtracks, which he analyses using six different parameters – which are not intended to provide a code, but as tools to establish some "*meaning potential*" (1999, p. 10, emphasis in the original) always to be referred to the specific context in which the sound event occurs. The parameters are the following:

1.   Perspective, which is determined by the relative loudness of simultaneous sounds and places sounds at different distances from the speaker;

2.   Time and rhythm, which refers to the tempo characterising sounds;

3.   Interaction of voices, i.e. how the plurality of voices involved in the soundscape intertwine (by turn-taking or simultaneously);

4.   Melody, which is realised through pitch movement, pitch range and pitch level, with pitch being modulated in order to convey specific feelings or emotions;

5.   Voice quality and timbre, which includes varying degrees of tension, roughness, breathiness, loudness, pitch register, vibrato;

6.   Modality, which refers to the degree of truth assigned to a sound and is determined by a number of factors, including pitch range, durational variation, variation in degree of loudness, and others.

In this study, we will focus on the parameter 'interaction of voices' by investigating how the plurality of voices involved in the audio guide soundscape interact to create meaning and involve the listener. This parameter was chosen because it can be considered to encompass many features relating to the other parameters. Indeed, the concept of 'voice' here is intended in both its strict and wider meaning, and more specifically 'voice' as 'speaker', on the one hand, with his/her generic or specific role within the narration, and 'voice' as 'semiotic resource', on the other hand, which includes sounds, music, and the way speech is prosodically delivered through voice (pauses, pitch, intensity, etc.). When relevant, other parameters will be referred to during the analysis.

As far as music, sounds and prosody are concerned, it should be made clear that the paper does not provide a systematic analysis of how often these occur and are used in the audio guides[3] because this would be outside the aim of this paper; instead, it provides numerous

----

3   For a full account on the use of music in city audio guides, see Fina (2017).

examples showing how the various voices may be combined to produce meaning and involve the visitor. On the basis of the theoretical premises outlined in this section, we will now define the research questions, describe the data and illustrate the methodology.

3. The corpus

The criteria for selecting the audio guides were chosen so as to make them fit all the levels of analysis illustrated in section 1. The study was carried out on a corpus of 50 professional city audio guides, grouped into three different groups as follows:

ITA corpus	Br+I corpus	USA corpus
17 Italian audio guides	15 British + 2 Irish audio guides	15 American audio guides + 1 audio guide of London by Rick Steves

Table 1. The three groups of audio guides

The audio guides were grouped so as to make the categories of 'language' and 'country' coincide by including audio guides of cities belonging to the same country in which they were produced. In this way, possible 'contaminations' in the producers' choices based on subjective perceptions of cities belonging to different countries have been avoided. The only exception is the audio guide of London produced by Rick Steves[4], but the inclusion of an audio guide of a British city produced by an American is not believed to significantly affect the research because this study does not aim to investigate the linguistic representations of tourist destinations in audio guides, nor the sociological factors underlying such representations. The audio guides were also selected so as to ensure diversification of producers, that is audio guides of the same city but created by different producers are included in the corpus, while audio guides of different cities created by the same producer do not occur. Following this criterion, only two Irish audio guides were retrieved on the Internet, and these were included in the same group since the two countries are geographically and linguistically close and, considering that the two Irish audio guides do not significantly differ from the British ones, excessive fragmentation would be superfluous.

On the basis of the theoretical premises illustrated in section 2, this paper sets out to investigate what strategies, at both the verbal and extra-verbal levels, are employed in the audio guides to make contents more enjoyable and, as a result, enhance visitor involvement. In order to answer this research question, the audio guides were qualitatively investigated focusing on two key elements: 1) the types of speaker that may be involved in narration and the different roles they have in delivering contents, and 2) the way speech, music, sounds, and prosodic features intertwine in relation to content and the expected effects that such

---

4   Rick Steves produces audio guides of European destinations only. See https://www.ricksteves.com/watch-read-listen/audio/audio-tours (last accessed February 2016).

combinations imply. Possible differences between the three groups of audio guides in the frequency of specific features will be highlighted and discussed and, at the end of the analysis, the results will be framed within a discussion on how the strategies adopted in the audio guides may fulfil public speaking principles, and then key pedagogical implications for both academics and professionals in the field will be highlighted.

4. Second speaker intervention and interaction of voices

The investigation of the voices interacting in the audio guides begins by looking at the number of speakers and the roles they may be involved in narration. To analyse this aspect, the number of audio guides featuring more than one speaker was identified in each group, as reported in Table 1:

ITA corpus	Br+I corpus	USA corpus
3/17	9/17	9/16

Table 2. Number of audio guides with more than one speaker

Interestingly, second speaker intervention is definitely more common in audio guides in English. Indeed, in both the Br+I and USA corpora, more than half of the audio guides feature more than one speaker, against only three occurrences in the ITA corpus.

A thorough analysis of the scripts showed that there may be different types of speakers, having different roles in the narration or dealing with specific type of content. As a matter of fact, the narrator may interact with other narrators, with local people (who may either be experts or ordinary people), and actors that impersonate fictional characters or lend their voices to report quotations. These roles are shown in Figure 2:
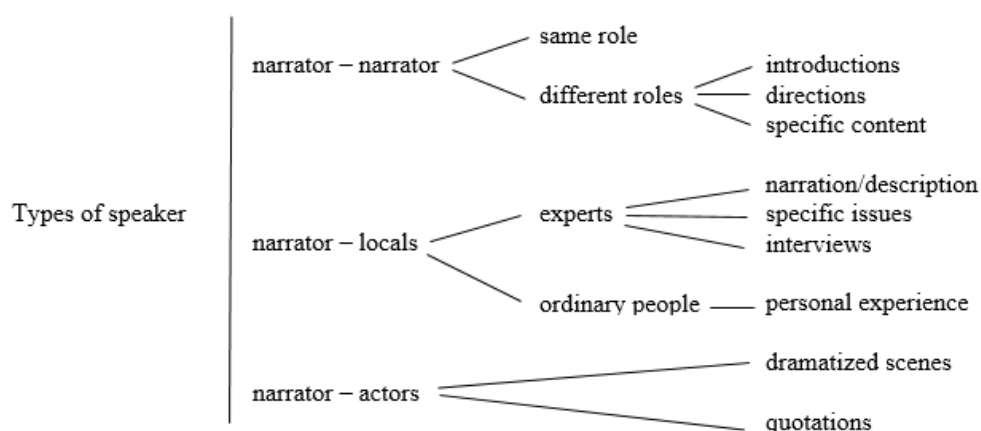

Figure 2. *Types and roles of speaker*

In the following sub-sections, we will discuss each role type.

4.1. Narrator – narrator

Table 2 shows the number of audio guides featuring two narrators in each corpus:

Table 3. *Number of audio guides with two narrators*

	ITA corpus	Br+I corpus	USA corpus
Narrator –narrator	3	6	7

Overall, in most of the audio guides with more than one speaker, the speakers are two narrators taking turns in narration. What we are interested in is whether the narrators interact while alternating each other or they simply deliver their own parts without referring to each other's presence. In the three Italian audio guides (*Modena*, *Perugia* and *Torino*), the two narrators simply alternate each other in narration or description without any specific criteria and without interacting with each other. In a number of audio guides in English, instead, the speakers have different roles, which are basically related to different content types, as can be seen in the following examples:

(1) *[female speaker]:* […] but it's here that your personal guide, the historian Eric Melvin, is waiting to tell you more.
[male speaker]: Hi, my name is Eric and it's my pleasure to be your guide for this tour of Edinburgh's Royal Mile. (Edinburgh Royal Mile, track 01, min. 01:12)

(2) *[female speaker]:* Now, head east and cross 11th Street. If driving, please turn on your emergency blinkers now, so others know you are going slowly. If walking, use the right side of the street. Stop by the corner of 11th and Main before listening to the next track.
[male speaker]: At 11th and Main, the building on the southeast corner was originally a silent movie theater […] (Lexington, min. 06:32)

(3) *[male speaker]:* To help you along, I've invited my colleague Lyssa. Welcome, Lyssa!
[female speaker]: "Hi"!
[male speaker]: She'll give directions from one stop to the next. (London by Rick Steves, min. 01:57)

In excerpt (1), the female speaker first briefly presents the tour and then introduces the guide that will provide information on the various sites. Throughout the audio guide, the female speaker also provides directions to the next stops. Similarly, in excerpts (2) and (3) the directions are provided by a female speaker, while the male speaker provides narration and description. Interestingly, while in excerpt (2) the switch between speakers and roles occur without verbal interaction, in excerpt (3) the two speakers create a dialogic dimension. It is worthwhile to further investigate this aspect, and we will focus on the following examples:

(4) *[female narrator]:* As you make your way to St. Clement Danes, listen to Rick give an overview of The City's history. Rick?
[male narrator]: This district, called "The City", stretches from near St. Clement Danes

to the Tower of London. (London by Rick Steves, min. 03:49)

(5) *[special narrator]:* Yo, GPS Gena! It's Philadelphia Frieda here! Before you finish the directions, make sure you tell them to stop at Liberty Boards.
[main narrator]: Why don't you tell them, Philadelphia Frieda, since it's another Philadelphia First?
[special narrator]: Ok. As you're walking along 7th Street, notice Liberty Boards Surf Shop on the right-hand side. So, why on earth have I stopped you guys in front of a surf shop? Did any of you know that the first flippers were invented by one of our Founding Fathers? During one summer in the early 1700s, Benjamin Franklin invented a swimming machine [...]. (Philadelphia Constitutional, track 05, min. 00:05)

It is evident, here, how the alternation in narration does not occur abruptly, but in a conversational way, with the female speaker in excerpt (4) signalling to the male narrator that he may start talking ("Rick?"). In excerpt (5), a special speaker – whose task is to illustrate the facts that happened for the first time in Philadelphia ('Philadelphia Firsts') – reminds the main narrator to show a Philadelphia Firsts-relevant item to the visitor; then the main narrator invites the special speaker to take the floor herself as that item is part of her area of competence; finally, the special speaker illustrates the 'Philadelphia First' related to Liberty Boards. The two rhetorical questions are part of the strategy as well, as they aim to catch the attention and raise curiosity in the visitor.

In the following excerpts, the roles of the speakers are not diversified according to content types, but the interaction between them is strategic because it is a way to provide the visitor with knowledge, or to explain specific concepts, without sounding didactic:

(6) *[female narrator]:* Now, walk south to the corner of North Pearl and Orange Streets and cross Orange Street to stand in front of the first church in Albany.
[male narrator]: You know Pat? Teddy Roosevelt prayed here. (Albany, min. 11:14)

(7) *[male narrator]:* He was impeached for malversation of office in the Commons, but was acquitted by the House of Lords.
[female narrator]: Malversation of office? What on earth is that?
[male narrator]: Well, the modern technical term for "malversation of office" is "fiddling the books". (Edinburgh New Town, track 2, min. 00:50)

In excerpt (6), the male narrator asks a rhetorical question to the female speaker as part of a strategy to introduce preliminary information about the church in a more interesting, less didactic way. Interestingly, this strategy occurs immediately after the female speaker has introduced the site, to raise the visitor's interest from the start, and the description does not start with historical details (year of foundation or similar), but with a particular aspect of it ("Teddy Roosevelt prayed here") that 'sacralises' the site and makes it unique. In excerpt (7), instead, the female speaker asks her colleague to clarify a concept. In a way, she plays the visitor's part by voicing a question that the visitor would probably ask if the tour were human-guided. This strategy occurs very frequently in *Edinburgh New Town*. Other examples are the following:

(8) *[male narrator]:* What do you mean "friend"?
[female narrator]: Hmm well, you know, someone she lived with.
[male narrator]: Ah, in Scotland we call that a "bidyin". (Edinburgh New Town, track 14, min. 01:53)

(9) *[female narrator]:* Which particular Edinburgh song had you in mind?
[male narrator]: Well, the one who sang the above, "Keep right on to the end of the road".
[female narrator]: Oh, so not the one who sang, "If you can say/"It's a braw bricht moonlicht nicht"/Then yer a'richt, ye ken."
[male narrator]: Yes, that's the one!
[female narrator]: Oh!
[male narrator]: Puzzled? All will be revealed at the end. (Edinburgh New Town, track 21, min. 01:43)

Here, interaction between speakers becomes a strategy not only to clarify concepts, but also to introduce aspects of the local language and culture and at the same time, so as to entertain visitors by proposing riddles that they have to puzzle over. In excerpt (9), even though it is made clear in advance that the song they are talking about is a traditional Scottish song ("Which particular Edinburgh song had you in mind?"), the speaker does not provide further details and raises the visitor's curiosity by saying that "all will be revealed at the end". It may be argued that the intention here is to leave the visitor with something unsolved that push him/her to go beyond the audio guide experience and look for more about the city and its culture.

The speakers may also interact to comment about themselves or what the previous speaker has said, or to joke:

(10) *[female narrator]:* This striking statue of the Emperor Trajan actually has nothing to do with London, since he never came here. But the local vicar put it in here for interest.
[male narrator]: Ah, thank you. However, look to your left, at the large pale stone plaque on the wall. (London by StrollOn, chapter 06, min. 00:24)

(11) *[female narrator]:* In the building itself is a portrait of George Washington in full Masonic regalia. The building was opened in 1859 as a Grand Lodge of Scotland. I would like to tell you more, but, as you know, it's a secret and I am but an unworthy woman.
[male narrator]: Ah, how true, how very true. (Edinburgh New Town, track 19, min. 01:34)

(12) *[female narrator]:* The walkway is lined with modern shops –
[male narrator]: Ah-ah, no shopping, no shopping!
[female narrator]: [laughing slightly] It eventually leads to the noisy traffic-filled boulevard called Cheapside. (London by Rick Steves, min. 44:34)

In excerpt (10) the male speaker thanks the female speaker for the information she provided, but it is not clear whether the thanking is genuine or sarcastic, considering that the male narrator immediately turns to a new topic. In excerpt (11), the female speaker expresses appreciation for the way her colleague delivers information, and in doing so she is probably trying to positively bias the visitor's own evaluation. In excerpt (12), instead, the male narrator jokingly exhorts the visitor not to deviate from the tour to do shopping ("Ah-ah, no shopping, no shopping!"), as the visitor might feel like doing after the female speaker made him/her notice that the street is lined with "modern shops". Amused by her partner's comment, the female speaker then continues to deliver information.

To conclude, a first key difference between audio guides in Italian and audio guides in English lies in the fact that, differently from the Italian ones, in the latter not only are the roles diversified in relation to content types (e.g. different speaker for directions), but the speakers may also interact to introduce specific concepts or to entertain the visitor with jokes and comments. Considering that these forms of interaction between narrators occur in-between more or less elaborate sections containing the description of the site, interactivity between narrators can be considered a strategy enacted to minimize the complexity of the written script – from which the narrator's speech originates – by inserting features of orality. These probably aim to keep the listener's attention alive, since s/he is very likely to notice the switch from a male voice to a female voice and vice versa, but also to cognitively reset the listener's mind by creating short breaks in the narration.

4.2. Narrator – locals

Second speaker intervention may also involve locals. This feature, though, characterises audio guides in English only, as can be seen in Table 4:

	ITA corpus	Br+I corpus	USA corpus
Narrator – locals	//	5	3

Table 4. Number of audio guides featuring local people intervention

The locals involved can either be experts or ordinary people with a direct tie or knowledge of a particular site. This type of second speaker intervention may occur following a structured sequence, according to which the narrator briefly introduces the site and the local expert describes it, as in the following excerpts:

(13) *[narrator]:* Take a little time to look around this large building that is made up of three sections, as Amy now takes us through its unusual development.
[local expert]: The King's Circus is the second phase of this great sequence of spaces that we have in Bath having started at Queen Square. [...] (Bath, track 05, min. 00:03)

(14) *[narrator]:* [...] you might like to take a stroll around the block and explore the exterior of the building, as Gordon Urquhart, architectural historian and conservator at the Glasgow City Heritage Trust, takes us inside. But, first, he explains how this

magnificent building came to be.

[local expert]: Glasgow City Chambers, or to give its official title of the Municipal Buildings of Glasgow, is a building that's so well-known, so well-loved and so prominent in our townscape that we often take it for granted. [...] (Glasgow, track 2, min. 00:11)

(15) *[narrator]:* It was designed as the city's war memorial as well as its concert hall. But what are all the columns about?

[local expert]: We can see another familiar feature, a row of columns, these are of the Corinthian style [...] (Sheffield, track 03, min. 00:15)

In *Downtown Houston*, instead, a local expert comes into play at a specific stop of the tour, i.e. when Christ Church Cathedral is introduced. As in the previous examples, the main narrator briefly introduces the site, and then an expert guide starts to describe the interior of the cathedral:

(16) *[narrator]:* Turn right and walk into the cathedral. Stand at the back for a moment, just take it all in. (Downtown Houston, track 10, min. 03:40)

[expert]: [amplified voice; religious song] Welcome to Christ Church Cathedral. I'm Judy Mood, and I'm your host and tour guide today. You are standing in the most historic building in Houston. Christ Church was the only religious institution in Houston for a while and so [...] (track 11, min. 00:00)

Here, soundscape comes into play as well. The expert's voice is amplified in order to reproduce the echo-like effect characterising high-ceilinged places, an effect that the visitor is missing at the time of visit because s/he is wearing headphones. Furthermore, the description is accompanied by a solemn religious song in the background, which has a mood-setting function[5] as its main aim is to strengthen the sacred atmosphere that the visitor is supposed to perceive while visiting the cathedral.

But expert intervention may also occur in an unstructured way, with the expert intervening at a given point during narration to add details to something previously said by the main narrator, as in *New York NAT*:

(17) *[narrator]:* As they sail, Hudson and his crew came face to face with the mysterious residents of Mannahatta, the Lenape Indians.

[expert 1]: At the entrance of the river they found the natives brave and warlike, but inside and up to the highest point of the river they found them friendly and civil, having an abundance of skins and furs.

[narrator]: To the Dutch, furs, particularly beaver pelts, were important items of trade.

[expert 2]: The beaver pelt was the first thing that they were interested in, which seems a strange and fantastic object to spend so much time over, but from that you

---

5   For an overview of the functions of music, see Cohen (1999).

could make felt, and felt was warm, and you would make very warm hats and these were prized all over Europe. (New York NAT, track 01, min. 08:14)

It is evident, here, how the expert's interventions are relevant to what the narrator said before, and occur to provide deeper insights into specific aspects. A possible effect of this type of narrator-expert alternation could be the fact that the visitor's knowledge is enriched in a way that sounds like an informal conversation rather than a formally delivered lecture on the topic.

Ordinary people with a strong tie with the site being described may also intervene to share their experience and knowledge:

(18) *[narrator]:* How cool is this? A perfect view of the ball field. Hear someone who can give you the real story on this place:
[local]: Welcome to Minute Maid Park, in magnificent Downtown Houston. I'm Milo Hamilton, voice of the Houston Astros. I started in 1950, so this is my 59th year doing baseball and we're very proud of this ball park [...] (Downtown Houston, track 07)

A particular example can be found in *Denver*, in which the local's intervention takes the shape of a real advertisement:

(19) *[narrator]: [jazz music and song]* In addition to being a Denver favorite, Lannie has worked with BB King, Ray Charles, Bill Cosby, Jay Leno and Roseanne Barr.
[local]: Hi, this is Lannie Garrett and I'd like to invite you to take a peek at our gorgeous little nightclub. [...] Our cover is a great venue for many different shows, from my swinging big band to my country comedy spoof, the Patsy DeCline Show. (Denver, stop 16, min. 03:26)

In this case the contribution from the local character is not merely informative, but also performs a persuasive function, as it aims to attract the visitor into the nightclub. The persuasive intent is enhanced by the musical soundtrack, which consists of a jazz song performed by the singer who is advertising the nightclub. Before ending the intervention, the local also provides contact information. Inserting this type of information in audio guides requires constant and accurate updating, since over time these places may also close or completely change.

The *Manchester* audio guide is entirely based on second speaker intervention. This tour was recorded on site, and is focused uniquely on the whole area along the Rochdale Canal. Far from describing physical places, this audio guide is meant to offer ten 'portraits' of the city of Manchester, which describe the role of the Rochdale Canal throughout history in the social, cultural, political and economic development of the city. These portraits are 'painted' through the opinions and experience of a great number of experts and locals alike. Descriptions in this tour do not concern architecture or directly observable features, but aspects of the city that could otherwise be learned only through study, research or life-long living in the area:

(20) *[narrator]:* So what's the lesbian scene like in Manchester, 'cause we hear a lot about the male gay community, but we don't hear much about the female gay community?
[local]: Oh, it's fantastic. Vanilla's been open now for nearly ten years and every year we're getting better, with more and more lesbians coming out in Manchester from all around the north west [...] (Manchester, track 9, min. 03:37)

(21) *[narrator]:* Are there any ground-breaking discoveries that were made by people from Manchester or passing through Manchester?
[expert]: Dalton produced the first atomic table as a result of which he's known as the father of modern chemistry. (Manchester, track 1, min. 02:14)

Contributions from locals may be considered valuable resources in the audio guide making not only because of the interactivity between speakers that they often require, but especially because they provide the visitor with interesting insights into the soul of the site and enrich the visitor's knowledge with information that is very unlikely to be found in any written guide. However, it should be remarked that these contributions are often delivered through spontaneous speech, which is significantly less controlled than speech delivered by the main narrator. Hence, difficulties in content uptake may arise due to features that normally characterise spontaneous speech, such as uncontrolled speech rate, syntactic inaccuracies, reformulations, and similar. The *Manchester* audio guide, for example, would be difficult to understand not only for non-native speakers of English but also to native speakers, due to the very fast speech rate, which is often combined with a very strong local accent. In *Glasgow*, interventions by experts are sometimes characterised by hesitations, repetitions, reformulations, and syntactic inaccuracies, which may require a higher cognitive effort and hence affect understanding.

4.3. Narrator – actor

This type of second speaker intervention occurs in a limited number of audio guides in English only, as reported in Table 5:

	ITA corpus	Br+I corpus	USA corpus
Narrator – actors	//	3	4

Table 5. Number of audio guides featuring actors

Narration may sometimes occur in the form of a play. In the Br+I corpus, the *Leicester* audio guide stands out for resorting to drama to narrate the historical events related to King Richard III's defeat at the Battle of Bosworth. This mode of narration involves the use of additional voices impersonating fictional characters, who are employed to give accounts of particular aspects of the historical events. The role of the main narrator, instead, is to briefly introduce the site and the related historical events first, and then the fictional characters and their roles in the story:

(22)   *[narrator]:* The battle is over. Rhys, a Welsh foot soldier in the army of the Stanley's, who fought for Henry, sits in the shade of an oak tree, on Ambion Hill.
[fictional character]: [raven's call] King Richard is dead. I can't believe it, even though my own eyes did see it. Even through the din we could see the battle was turning against Richard. And then he charged. [soldiers' screams begin] Him and all his knights. Heading for Henry he was. Looking to end things quick I reckon. We thought he'd done it too. When Henry's standard fell we thought that was it, but his men rallied *[soldiers' screams].* The charge broke against the wall of pikes and many of Richard's knights were unhorsed *[horse cries].* (Leicester, track 06)

This mode of narration, which can be classified as 'faction' (i.e. facts + fiction), is adopted to bring to life historical events, or narrate particular aspects of them, through the words of fictional characters. Thus, this is a form of *edutainement*, by which history is 'taught' to the visitor by entertaining him/her with dramatized scenes, whose aim is to conjure vivid pictures in the visitor's mind. To do so, cinematic techniques and prosody are used accordingly. Environmental sounds reproducing the natural setting allow the visitor to picture the setting of the scene, with the raven's call making this setting particularly menacing. Furthermore, sounds reproducing soldiers' screams and horse cries in the culminating moments of narration evoke the harshness and confusion of the battle field. At the prosodic level, the parameters 'Tempo' and 'Voice quality' come into play, since the actor makes two distinct pauses before uttering "died" and "he charged" to stress the importance of the facts and to load them with gravity in the former and anticipation in the latter; while the emotional load of the scene is expressed in the tension and roughness of his voice when he utters "he charged". It is not by chance that this narration technique is employed in audio tracks regarding sites fit for narration but not for description (e.g. Bow Bridge, Ambion Hill), where there is nothing in particular to look at and a conventional narration revolving around a succession of historical facts might disperse the visitor's attention after a while. However, the dramatization of history may also determine problems in content uptake. These are related to linguistic issues, because the actors involved speak in local accent, and idioms and particular expressions are occasionally used. Although these features might affect understanding by non-native speakers of English, this risk is minimised by the fairly slow pace of speech.

The *Edinburgh Royal Mile* tour is particularly dense with dramatized scenes. These, however, usually consist of one utterance only and occur when narrating particular episodes about characters, as in the following excerpt:

(23) *[narrator]:* In 1637, though, King Charles I was determined to force the Church of Scotland to become like the Church of England, with the King as head, supported by bishops and archbishops. It was announced that St Giles was now to be a cathedral. Bishops and an archbishop were reintroduced. This was too much for the Scots. A new prayer book was introduced, which provoked a riot. *[noises of the protesting crowd start]* Amongst the congregation that day was Jenny Geddes, a street-seller of herbs, who was sitting on her stool to hear the service. Feelings were running high. When the Dean started to read from the prayer book, she stood up and shouted,

---

6   'Dare you say the Mass in my ear?'

> *[actor]:* 'DAUR YE SAY MASS IN MY LUG?[6] TAKE THIS FOR HIS MAJESTY!'
> *[narrator]:* An enraged Jenny Geddes flung her stool at the Dean *[noise of stool thrown]*, narrowly missing his head. There was uproar. The service was abandoned. (Edinburgh Royal Mile, stop 6, min. 06:47)

The dramatized scene refers to an episode happened in the Church of St Giles and framed within the wider historical context of the conflict between the Church of England and the Church of Scotland. The narration of the event would be equally informative without the dramatized part, but this was introduced to allow the visitor to picture the event in his/her mind, thus making the event itself more interesting and memorable. Again, vivid pictures of the scene are conjured in the visitor's mind by the use of sounds reproducing people shouting in protest, and the noise of the stool being thrown at the Dean. At the prosodic level, a high voice intensity (indicated with capital letters) is employed to convey the characters' emotional load. However, local language and accent might limit understanding by non-native speakers of English, although the gist of the utterance can be deduced from the context.

The following example is particularly interesting because it features a 'play within the play':

> *(24) [narrator]:* If you were walking here in September 1665, you'd hear cries of 'Bring out yer dead! Bring out yer dead!'
> *[actor]:* 'But I'm not dead yet!'.
> *[narrator]:* It was men patrolling the streets with wagons to cart away dead bodies.
> *[actor]:* 'But I'm not dead!'.
> *[narrator]:* Shut up! (London by Rick Steves, stop 14, min. 30:46)

The exclamation 'Bring out yer dead' was used during the time of the Black Plague to let people know that the cart for the dead bodies was passing by. The answer "But I'm not dead yet!", uttered by another voice, is a quote from the 1975 surreal comedy *Monty Python and the Holy Grail*, written, directed and performed by the comedy group Monty Python (or 'The Pythons'). The quote refers to the scene in which the cart collecting corpses is passing by and a particularly resilient victim insists on not being dead.[7] The main speaker, i.e. Rick Steves, intervenes in the scene by hushing the insisting victim. What is interesting about this narrative mode is the fact that it occurs not at the end, but at the beginning of the historical account of the Great Plague in London. Thus, the visitor's attention is drawn on what could be considered an unpleasant topic by referring first to a particular detail of the event ("Bring out yer dead!") and then to popular culture (the quote from the *Monty Python* film). This reference, though, might not be caught by visitors belonging to different cultures, or by young generations of visitors. In such case, the visitor would only catch the surface part of humour, as s/he will not be able to fully understand its intertextual reference.

Actors may also be involved to report archival material, i.e. information contained in any type of documents recording events, such as written accounts, unpublished letters, excerpts from magazines and similar. The main aim is to enrich the visitor's knowledge of specific aspects of historical events through authentic sources that would be otherwise inaccessible for visitors, as in the following excerpts:

---

7   The scene can be viewed at https://www.youtube.com/watch?v=grbSQ6O6kbs (last accessed March 2016).

(25) *[narrator]:* Using the Arkansas River, many passed through Little Rock, particularly the Cherokee. In 1830, the *Arkansas Gazette* carried this news item:

[actor]: "Captain Johnson arrived at this place about noon on Wednesday last, having on board about 100 cabin and deck passengers, principally emigrants to the Territory, and about 200 emigrating Cherokee Indians, from the old nation, who are on their way to the Cherokee Country up the Arkansas." (Little Rock Arkansas, track 02, min. 02:26)

(26) *[narrator]:* The 4ᵗʰ division formed up on Main Street for about a half mile to the east. A Lieutenant, named Barlow, described the scene before the battle:
[actor]: "the street was lined with ladies, sobbing and waving handkerchiefs, and one old gentleman climbed a gate post and exhorted our marching column in true and camping style. General Price had surely posted his brass band under cover at the foot of the hill and as we approached they struck up a lively battle tune, 'Garryowen'". (Lexington, min. 04:21)

In excerpt (25), the actor reports the piece of news from the *Arkansas Gazette* using his voice in the same way as a TV news journalist or radio speaker would do, i.e. modulating pitch range to keep the listener's attention and using pitch accent on specific words to give rhythm to the narration. In this way, the information is made more authentic. Prosody plays an important role also in excerpt (26), with the actor's voice stressing the words "exhorted", "lively" and "Garryowen" to give emphasis to the narration.

Finally, in *Philadelphia Constitutional* excerpts from presidential speeches are reproduced by an actor imitating the president's voice, accompanied by solemn music:

(27) *[narrator]: [solemn music]* It was here Thomas Jefferson wrote the Declaration of Independence, in June of 1776. At that time, this house was considered on the outskirts of town.
[actor]: "When in the course of human events it becomes necessary for one people to dissolve the political bands which have connected them with another [...]." (Philadelphia Constitutional, track 04, min. 00:14)

A soft amplifying effect has been applied to the actor's voice, so as to reproduce the original aural setting of the event and make the visitor feel as if s/he were truly attending the event, while the music matches the theme of independence and the solemnity of Jefferson's speech.

In this section, we have highlighted another important difference between audio guides in Italian and audio guides in English, that is the fact that dramatized scenes, exactly like contributions from locals, characterize only a number of audio guides in English. The importance of dramatized scenes lies in the fact that they make way for soundscape to come into play, with ambient sounds and/or music intertwining with contents so as to create vivid pictures in the visitor's mind, and prosodic features of speech such as pauses, pitch, and intensity being modulated in order to convey the emotional load. This can be considered a strategy to

reduce the complexity of the script by making contents more enjoyable, and consequently more memorable, provided that features that might compromise understanding, such as local accent, too fast speech rate, or the use of idiomatic language, are taken into account and reduced to the minimum in advance. At the practical level, for example, audio guide producers wishing to insert in the script interviews to local experts should recommend using plain language and controlling speech as much as possible. In the next section, we will see how different voices may also be produced by the same speaker.

4.3.1. Footing

In the previous section, we have provided examples of actors involved in audio guides to dramatize historical events or to report archival material. But the dramatized part or reported information may also be uttered by the same speaker. In such cases, the speaker momentarily abandons his narrator role and becomes an actor who lends his/her voice to report someone else's words.

This narrator-actor role shift can be defined as a change in "footing" (Goffman, 1981, p. 128), that is a shift in the speaker's projection towards an utterance, which involves sound markers such as pitch, volume, rhythm etc. Therefore, this aspect as well may be classified as a form of interaction of different voices. No striking quantitative difference in the number of audio guides could be detected, as shown in Table 6:

	ITA corpus	Br+I corpus	USA corpus
Footing	7	4	5

Table 6. Footing

This is no surprising, as the presence of second speaker intervention is more frequent in audio guides in English against the tendency of Italian audio guides towards one speaker only (see Table 3). Examples of footing from the corpus are the following:

(28) When the Lord Mayor was woken to see it *[noise of flames]*, he grumpily observed, "Augh! A woman might piss it out", and went back to bed – not his finest moment. (London by StrollOn, chapter 15, min. 02:01)

(29) And every night the world's oldest military ceremony takes place at the Tower of London – that started back in the 14th century. One group of warders secures the Tower and is then halted on their return by a sentry who challenges them:
- 'Who goes there?'
- 'The keys!'
- 'Whose keys?'
- 'Queen Elizabeth's keys.'
- 'Pass Queen Elizabeth's keys. All is well.' (London by StrollOn, chapter 8, min. 02:25)

In excerpt (28), the speaker abandons for a few seconds his role of narrator to impersonate the Lord Mayor, by adjusting voice pitch and intensity to provide a faithful representation of the episode. His tone becomes scornful, and his voice is rough when the exclamation "Augh!" is uttered. Furthermore, sounds reproducing flames in the background make the narration more vivid. In excerpt (29), the Ceremony of the Keys is 'staged' through the speaker's voice. This ceremony is difficult to access because attendance has to be booked long in advance. Thus, the reproduction of the ceremony in the audio guide aims to create a picture of the event in the visitor's mind, allowing him/her to virtually access the ceremony. The speaker's voice becomes solemn and authoritative, but no ambient sounds are used to enhance the narration.

Footing occurs not only in audio guides in English but also in a few Italian ones:

(30) Un articolo del comandante Augusto Meriggioli ne dà un resoconto dettagliato:
*[*The Firebird – Infernal Dance Of All Kastchei's Subjects *by Igor Stravinsky begins]*
"Era il 9 aprile 1970. Sul porto di Genova si era abbattuto un fortunale, con libeccio forza 7-8 e mare grossissimo. [...][8] (Genova, track 1.2, min. 04:00)

(31) [...] è ancora Dante a raccontare che già in vita Giotto era conosciuto come eccelso pittore, e che al tempo in cui dipinse la Cappella degli Scrovegni aveva ormai superato in abilità il suo maestro, Cimabue:
"Credette Cimabue nella pittura
Tener lo campo, ed ora ha Giotto il grido,
sì che la fama di colui è scura."[9] (Padova, track 1-1, min. 04:53)

In excerpt (30), the narration of the sinking of the *London Valour* through personal accounts rather than through the speaker's own words is meant to have an emotional impact on the listener. The detailed description of the tragedy, accompanied by powerful, pressing orchestra music reflecting the gravity of the event, once again allows the visitor to display a vivid picture of the tragedy in his/her own mind. When listening to the excerpt we immediately realise that the music accompanying this content, i.e. the *Infernal Dance* by Stravinsky, is perfectly in line with the tragic event that is being narrated. In particular, 'Tempo and rhythm', 'Melody' and 'Modality' play a key role in co-building the narrative. The destructive force of the sea is matched by the raging rhythm of the music; the sense of danger is expressed by the low-pitched voices of the wind instruments, alternated by the high-pitched voices of the string instruments, whose high loudness heightens the tension. Perceptually, the intensity range between music and speech is narrow: this adds further value to the narrative, but at the same time does not affect speech, which is still clearly audible. In excerpt (31), instead, the same

---

8   A detailed report is provided in an article by Chief Augusto Meriggioli: "It was 9 April 1970. A storm had fallen upon the Genoa harbour, with the wind gusting hard and the sea waves getting bigger and bigger. [...]"

9   It is once again Dante telling us that Giotto was considered an excellent painter also in life, and that by the time he painted the Scrovegni Chapel his skills had already overcome his master Cimabue's:
"Less bright succeed not. Cimabue thought
To lord it over painting's field; and now
The cry is Giotto's, and his name eclipsed."
(Charles, W. E. (ed.). (1909). *The Divine Comedy of Dante Alighieri, Hell, Purgatory, Paradise.* The Harvard Classics.)

narrator recites a few lines from Dante's *Purgatory*. Again, prosody plays an important role, with the speaker adapting his voice to mark the distinction between the narrated part and the quoted part. When he utters the quotation, his voice becomes solemn and the rhythm slows off on the passage "*ha Giotto il grido*", while a pause before "*è scura*" confers emphasis to the concept.

All these examples show that interaction of voices in audio guides occurs also when two roles are performed by the same voice. Indeed, footing implies two distinct roles and hence two distinct voices, and prosody plays a fundamental role in conveying this distinction to the listener.

5. Discussion

In most cases, the audio guide as a text genre is a written script delivered orally, and in this study we raised issues about the need to adopt specific strategies to make such written text fit for oral delivery, so as to make contents more enjoyable and capture the visitor's attention. The analysis shows that such strategies are visible in narration modes that involve interaction between different speakers and interaction between speech, music and sounds. More specifically, the dialogic dimension created by two narrators fits the orality of the genre, as it creates breaks in an oral text which is dense with historical information, and constitutes a strategy to involve the visitor by means of comments or jokes between the speakers, or to enrich the visitor's knowledge by means of questions and answers designed to introduce contents or explain concepts. Going back to genre hybridization as discussed in section 2.2, the interdiscursivity process enacted in a number of audio guides is 'genre embedding', since contributions from experts and interviews to locals are features characterising the documentary genre, while dramatized scenes, archival material and quotations, as well as the interaction of speech with music and sounds are features belonging to the film/theatre genre. These instances of genre embedding allow us to shed light on three possible subgenres of the audio guide genre, i.e. 'audio-documentary', which seems to be particularly prominent in American audio guides, with contributions from locals or insertion of archival material; 'audio-faction' (i.e. facts and fiction), which mainly characterises the Br+I corpus, with the use of drama and cinematic techniques to bring to life episodes that truly happened in the past, and 'live recorded audio guide', such as the *Manchester* audio guide, which is entirely based on interviews recorded on site.

In the light of the findings of this study, the audio guide genre can now be framed within public speaking, as it shares similarities with the public presentation genre. One of the principles of public speaking provided by Griffin (2014) – and in the existing literature in general – relates to resorting to specific narration techniques to catch the attention, clarify concepts, reinforce points, bring concepts to life by evoking pictures in the audience's mind, and impact the audience's emotions. Contributions from experts and interviews to locals have the potential to make the audio guide more interesting and appealing, as they are meant to provide the visitor with deeper insights into the history of the sites and their cultural significance, by delivering information that the visitor is unlikely to find in ordinary guides. Similarly, dramatized scenes, archival material and quotations are meant to impress specific aspects of the site in the visitor's mind by bringing to life historical episodes or particular aspects of

them, and by voicing not only the characters involved but also their emotions and the whole atmosphere characterising the scene, through music and/or sound effects reproducing the aural setting of the scene. We have also seen that prosody plays a significant role, as it contributes to conveying emotions, especially when pauses are used before specific words so as to emphasise them or load them with emotional meanings, in line with the principle of public speaking according to which intentional pauses should be used to emphasise concepts.

Finally, it is worth highlighting that most of the strategies identified in the analysis tend to occur in audio guides in English only, and although we are dealing with small numbers, questions arise about whether the audio guide genre is conceived differently across different languages and cultures. Such questions, however, can only be answered through further investigation not only of the genre itself but also of the reasons underlying audio guide producers' choices.

6. Pedagogical implications

The pedagogical implications of this study mainly revolve around producing writing for oral delivery. The analysis carried out in this paper shows that the audio guide script, in order to be made fit for (effective) oral delivery, may undergo hybridization processes by which features of other genres are inserted in the audio guide in order to make the contents more appealing. More specifically, we may argue that producing audio guide scripts requires competences in creative writing, so as to make the complex informative contents of the script more appealing by means of *ad hoc* verbal and extra-verbal strategies. This issue opens up multi-faceted pedagogical perspectives for academic and professional contexts alike.

First of all, studying the strategies identified in section 4 in an ESP class will trigger the creativity of students, who may in turn practise creative writing for oral delivery on 'aseptic' informative tourist texts and challenge themselves in turning complex texts dense with historical information into appealing audio guide tracks, provided that they hold written English competences at least at B2/C1 levels of the *Common European Framework of Reference for Languages*. On a more specific level, students of translation between Italian and English might find the differences between audio guides in the two languages interesting, as questions arise about how to deal with the English translation of Italian audio guides that feature none of the strategies identified in this study. Thus, students will acquire insights about cross-cultural implications and will start dealing with the practice of translation becoming aware of the function of processes like adaptation, re-writing, and even 'transcreation'. Furthermore, the differences found between Italian and English encourage further cross-language investigation of the audio guide genre involving other languages, in order to determine whether the genre is conceived differently across different cultures. The findings may also encourage students and academics interested in 'social semiotics' (Hodge and Kress 1998; van Leeuwen 2005) to further explore the audio guide genre, as well as similar genres, as a multimodal communicative artefact, in order to unveil strategies for meaning making.

Furthermore, the findings of this study may also be of interest to trainers and professionals for teaching speaking skills to students or trainees wishing to pursue a career in tour guiding, since they might take interesting insights from the various narration techniques, including the use of prosody to mark specific segments of information.

Obviously, professionals already working in the field of communication for cultural heritage promotion will benefit from training programmes in this area, since the proliferation of tourist promotional material – especially in English as a lingua franca – is not always paralleled by innovation and attention to users' needs.

References

Alfaro, I., Nardon, M., Pianesi, F., Stock, O., & Zancanaro, M. (2005). Using cinematic techniques on mobile devices for cultural tourism. *Information Technology & Tourism*, 7, 61-71.

Antelmi, D., Santulli, F. & Held, G. (2007). *Pragmatica della comunicazione turistica*. Rome: Editori Riuniti.

Bhatia, V. K. (1997). Genre mixing in academic introductions. *English for specific purposes*, *16*(3), 181-195.

Bhatia, V. K. (2010). Interdiscursivity in professional discourse. *Discourse & Communication*, *4*(1), 32-50.

Charles, W. E. (Ed.) (1909). *The Divine Comedy of Dante Alighieri, Hell, Purgatory, Paradise.* The Harvard Classics.

Cohen, A. J. (1999). The functions of music in multimedia: A cognitive approach. In S. W. Yi (Ed.), *Music, mind, and science*, (pp. 53-69). Seoul, Korea: Seoul National University Press.

Crystal, D. & Davy, D. (1969). *Investigating English style.* London: Longman.

Crystal, D. (1994). Refining stylistic discourse categories. In G. Melchers & B. Warren (Eds.), *English linguistics in honour of Magnus Ljung*, (pp. 35-46). Stockholm: Almqvist & Wiksell.

Denti, O. (2007). A multimodal investigation of tourist texts and cityscapes. In L. Jottini, G. Del Lungo & J. Douthwaite (Eds.), *Cityscapes: Islands of the Self*, (pp. 327-347). CUEC: Cagliari.

Denti, O. (2015). Gazing at Italy from the East: a multimodal analysis of Malaysian tourist blogs. *Lingue, Culture, Mediazioni*, *2*(15), 47-68.

Fina, M. E. (2016). *On Effective Audio Guiding: A Multimodal Investigation of Italian, British and American Audio Guides* (Unpublished PhD thesis). University of Salento, Lecce, Italy.

Fina, M. E. (in press). Prosody in tourist oral communication. Pausing in city audio guides. *Lingue e Linguaggi, 23* ESE Publications.

Fina, M. E. (2017). The role of music in city audio guides in Italian and in English. *Multimodal Communication*, *6*(2), 87-107.

Francesconi, S. (2011a). Images and writing in tourist brochures. *Journal of Tourism and Cultural Change*, *9*(4), 341-356.

Francesconi, S. (2011b). Multimodally expressed humour shaping Scottishness in Tourist Postcards. *The Journal of Tourism and Cultural Change*, *9*(1), 1-17.

Francesconi, S. (2011c). New Zealand as 'the youngest country on earth': a multimodal analysis of a tourist video. *Textus*, *24*(2), 323-340.

Francesconi, S. (2014). *Reading Tourism Texts: A Multimodal Analysis.* Bristol: Channel View Publications.

Francesconi, S. (2015). 'Dreaming in Italy': ibridazione e creatività in un digital diary video. In M. Rocca Longo & M. Pennacchia (Eds.), *Turismo creativo e identità culturale*, (pp. 45-57). Roma: Roma TrE–Press.

Francesconi, S. (2016). Green hypermodality: a social semiotic multimodal analysis of the 'Green your house' webpage. In O. Palusci (Ed.), (pp. 251-262). Green Canada, Brussels: Peter Lang.

Francesconi, S. (2017). Digital travel videos as ways of visiting basilicata: a multimodal genre analysis. In M. Gotti, S. Maci & M. Sala, (Eds.), *Ways of Seeing, Ways of Being: Representing the Voices of Tourism*, (pp. 49-71). Bern: Peter Lang.

Gebbensleben, S., Dittmann, J. & Vielhauer, C. (2006). Multimodal audio guide for museums and exhibitions. *Proceedings of SPIE Electronic Imaging – Multimedia on Mobile Devices II*, 6074.

Goffman, E. (1981). *Forms of Talk*. Philadelphia: University of Pennsylvania Press.

Griffin, C. L. (2014). *Invitation to Public Speaking.* 4th Ed. Cengage Learning and National Geographic.

Hodge, B. and Kress, G. (1998). *Social semiotics Opens in a new window*. London: Polity Press.

Holt, S. (2007). How to make an audio tour. Ten tips from Audissey Guides pioneer Rob Pyles. *Matdor Network* blog. Retrieved fromhttp://matadornetwork.com/notebook/how-to-make-an-audio-tour-ten-tips-from-audissey-guide-pioneer-rob-pyles/ (last accessed February 2016).

Ling Ip, J. Y. (2008). Analyzing tourism discourse: A case Study of a Hong Kong travel brochure. *LCom Papers*, *1*, 1-19.

Luyten, K. & Coninx, K. (2004). "Imogl: take control over a context aware electronic mobile guide for museums". *Procedings of the International Conference on HCI in Mobile Guides.*

Maci, S. (2007). Virtual touring: The web–language of tourism. *Linguistica e Filologia*, *25*, 41-65.

Maci S. (2013). *Tourism Discourse: Professional, Promotional and Digital voices.* Genova: ECIG.

Manca, E. (2016). Official tourist websites and the cultural communication grammar model: Analysing language, visuals and cultural features. *Cultus*, *9*(1), 2-22.

Neves, J. (2012). Multi–sensory approaches to (audio) describing visual art. In R. Agost, P. Orero & E. Di Giovanni (Eds.), (pp. 277-294), *Multidisciplinarity in audiovisual translation*, special issue of *MonTi*, *4*.

Neves, J. (2016). Enriched Descriptive Guides: a case for collaborative meaning-making in museums. In D. Katan & C. Spinzi (Eds.), *Tourism Across Cultures. Accessibility in Tourist Communication*, special issue of *Cultus*, *9*(2), 137-154.

Petrelli, D. and Not, E. (2005). User-centred design of flexible hypermedia for a mobile guide: reflections on the hyperaudio experience. *User Modeling and User–Adapted Interaction*, *15*(3-4), 303-338.

Santulli, F. (2007). The rhetoric of multimodality: The persuasive power of verbal-visual interplay. In G. Garzone, P. Catenaccio & G. Poncini (Eds.), *Multimodality in Corporate Communication: Webgenres and Discursive Identity*, (pp. 31-54). Milan: Franco Angeli.

Tallon, L. (2006). On Audio Tours: An Unknown Quantity. *Engage, 18*, Retrieved from http://www.engage.org/downloads/1AA029E3_18.%20Lo%C3%AFc%20Tallon.pdf (last accessed February 2016)

van Leeuwen, T. (1999). *Speech, Music, Sounds.* London: Macmillan.

van Leeuwen, T. (2005). *Introducing Social Semiotics*. London: Routledge.

Websites

Quirk, R. (1959). *The Survey of English Usage* (Quirk Randolph. 1959)

<http://www.ucl.ac.uk/english–usage/> (last accessed February 2016)

Monty Python – Bring out your dead!

<https://www.youtube.com/watch?v=grbSQ6O6kbs> (last accessed February 2015)

Websites of the audio guides cited (all last accessed February 2015)

Genova <http://www.galatamuseodelmare.it/cms/mirador-11837.html>

Padova <http://www.padovanet.it/informazione/i-padova-percorsi-turistici-audioguidati it>

Edinburgh Royal Mile <www.realhistorytalks.com>

Edinburgh New Town <edinburghtour.com/Uncle Bob>

Glasgow <walkingheads.bandcamp.com>

Leicester <leicester.audio-guide.co.uk>

London by StrollOn <www.strollon.co.uk>

Manchester <www.hearmanchester.com>

Sheffield <https://www.sheffield.ac.uk/archaeology/community-outreach/neoclassical

Albany <http://www.albany.org/visitors-center/walking-tour/>

Downtown Houston <http://www.audisseyguides.com/houston-downtown-2/>

Lexington <http://www.visitlexingtonmo.com/audio-tours.html>

Little Rock Arkansas <http://www.littlerock.com/things-to-do/audiotour/>

London by Rick Steves <https://www.ricksteves.com/watch-read-listen/audio/audio-tours/london>

New York NAT <http://nyharborparks.org/visit/tour-new-amsterdam-self-guided.html>

Philadelphia Constitutional <http://www.theconstitutional.com/audio-tours/audio-tours>