Multimedia Analytics Challenges and Opportunities for Creating Interactive Radio Content

Peer-reviewed author version

Bailer, Werner; WIJNANTS, Maarten; LIEVENS, Hendrik & Claes, Sandy (2020) Multimedia Analytics Challenges and Opportunities for Creating Interactive Radio Content. In: Lecture Notes in Computer Science, 11962, p. 375 -387.

DOI: 10.1007/978-3-030-37734-2_31
Handle: http://hdl.handle.net/1942/31479

Multimedia Analytics Challenges and Opportunities for Creating Interactive Radio Content

Werner Bailer[1], Maarten Wijnants[2], Hendrik Lievens[2], and Sandy Claes[3]

[1] JOANNEUM RESEARCH, Graz, Austria, werner.bailer@joanneum.at
[2] Hasselt University – tUL – EDM, Diepenbeek, Belgium
{maarten.wijnants,hendrik.lievens}@uhasselt.be
[3] VRT, Brussels, Belgium, sandy.claes@vrt.be

Abstract. The emergence of audio streaming services and evolved listening habits notwithstanding, broadcast radio is still a popular medium that plays an important role in contemporary users' media consumption mix. While radio's strength has traditionally lain in the shared live experience that it enables, radio shows are nowadays much more than just a linear broadcast feed: they are about user engagement and interaction over a amalgam of communication channels. Producing interactive radio programs in a cost-effective way thus requires understanding and indexing a large set of multimedia content, not just audio data (but also e.g., text, user-generated content, web resources). This position paper discusses a number of open challenges and opportunities with respect to the application of multimedia analytics in interactive radio production that need to be addressed in order to facilitate content creation at scale. Our work aims to support the future evolution of (interactive) radio, this way helping the radio medium to stay relevant in an ever-changing media ecosystem.

Keywords: radio production, interactivity, automatic content annotation, content structuring, summarization, context, adaptation, personalization, user-generated content

1 Introduction

The media landscape is undergoing significant changes, caused by the increasingly diverse and personalized ways in which media is consumed (in terms of time, place, context and content). These changes also impact radio, a medium that has maintained its popularity over a century, despite fierce competition. The proportion of Europeans who listen to the radio at least once a week is stable at 75%, the number of daily consumers is at 50% [1]. While the primary strength of broadcast radio is still the fact that it offers a shared live experience to listeners, radio shows are nowadays much more than merely a linear broadcast feed. Indeed, radio increasingly revolves around user engagement and interaction over a multitude of channels, including direct communication (phone, email, instant messaging), social media and visual radio. Many of the devices used to access these channels have screens and provide rich interaction capabilities. For example, 52% of UK adults have downloaded a radio app and 8% of UK adults regularly use radio catch-up services [34]. Radio is thus a medium embedded in a context of social media,

interaction and personalization. As such, radio is no longer only about audio and music, but increasingly also about the text, image and video content that is exchanged as part of the listener interactions in contemporary radio consumption practices. This observation has direct implications for radio production workflows, where creating appealing supplemental material and interactive content has become a necessity.

This position paper contributes a number of open challenges and opportunities with respect to the application of multimedia analytics in radio production. We hereby focus specifically on the authoring of interactive instead of purely linear radio content. The identified open challenges are grounded on empiric evidence harvested via desk research, the authors' personal experience in the radio research domain and, perhaps most importantly, the insights from discussions at the Workshop on Interactive Radio Experiences (IRE) at ACM TVX 2019 [9]. The IRE workshop was attended by both practitioners (e.g., radio producers) from broadcasters and researchers from broadcasters' R&D labs, which resulted in a wealth of "insider information" about (interactive) radio production and its practical pitfalls. Identified open challenges are first described on a high level in the remainder of this section (see Section 1.1 and Section 1.2), before diving deeper into the specific challenges and opportunities in terms of *content enrichment*, *context derivation and adaptation* and *production user interfaces* (in Section 2-4, respectively). A secondary contribution of our work is that it helps uncovering research gaps in the audio domain that typically have already been solved for other media like video and text.

1.1 Challenges for content production

The creation of interactive radio experiences is still in an early stage, but very interesting content has already been realized. For example, the BBC has created interactive audio content for smart speakers such as "The Inspection Chamber"[4] or "The Unfortunates"[5]. Listeners can influence the continuation of the story with their interactions, resulting in a myriad of possible paths through the narrative. Similar opportunities have also been explored outside the broadcasting domain. For example, the Financial Times produced interactive audio city guides for Google Assistant[6]. Another dimension of interactivity with audio content is introduced by object-based audio [36], which captures the scene as a set of audio objects and associated metadata, which can then be dynamically composed to a customized stream for individual listeners as needed.

However, at both public and commercial broadcasters, radio content production faces budget limitations. While creating rich and bespoke interactive content, built around a hand-crafted story, is likely to provide the most engaging experience, this approach is not economically viable at scale. In addition, there is considerable risk how new forms of content will be received by the audience. The alternative is thus to build on existing content, which has an installed consumer base, and to augment that content with supplemental material and by providing interaction capabilities. An approach as described by Baume, which uses existing radio shows to create enriched

[4] `https://www.bbc.co.uk/taster/pilots/inspection-chamber`
[5] `https://www.bbc.co.uk/taster/pilots/unfortunates`
[6] `https://www.ft.com/hiddenberlin`

and interactive podcasts [6], is thus a more realistic deployment example for the near future. Despite the more incremental nature of interactive content that is produced in this way, the required additional information, links, structure, etc. is currently typically still added manually. Richer time-based metadata is required to streamline and automate this process.

Interactive radio content is consumable on many platforms, including digital radio receivers, smart phones, tablets, smart speakers and TVs. The presentation (e.g., display size) and interaction capabilities (e.g., touch screen, voice interaction) are different for each of them. Interaction options will hence need to be presented in different ways, and must support different means to respond to it. Production tools must be able to simulate this range of options, support authors in defining interaction alternatives for different platforms, and enable them to understand how interaction options will be used and which issues may arise.

Listener co-creation processes grounded on user-generated content (UGC) are one particularly cost-effective way to enrich radio shows and to stimulate interaction. An example case for using UGC in event-related radio content is described in [14]. As UGC-driven co-creation commonly leads to extensive pools of content, the production team needs algorithmic support to (semi-)automatically select contributions that are of good quality in both technical and content aspects.

Finally, fine-grained audience measurements become a necessity in interactive radio experiences, as content creators must become aware how interactive data that supplements the radio signal is used and consumed, and how listeners interact with the radio program. Also, as the range of possible consumption contexts is wide, content should be able to adapt to user context (semi-)automatically at least to some degree. This adaptivity requirement again adds complexity to the radio production process.

1.2 A case for multimedia analytics

Many of the discussed production challenges have in common that a solution can be enabled by fine-grained metadata about professional content and UGC as well as detailed audience measurements. The extraction of semantically meaningful information about the content requires state of the art natural language processing and computer vision tools, which in turn yields large amounts of data that need automation for structuring and organization. However, what sets traditional media apart from the pervasive amounts of content on the web is the curation and creation of a narrative by human creators. Thus the final creative choices and decisions need to stay in the hands of the radio creators, supported by intelligent tools. The requirements for handling large amounts of heterogeneous multimedia data in order to enable creators to efficiently produce richer content makes this a multimedia analytics problem. Despite this, very few papers dealing with multimedia analytics actually address radio content (beyond mentioning the word radio in the introduction); see, for example, [26] and [10].

2 Content Enrichment

The first area related to multimedia analytics for the authoring of interactive radio content concerns enrichment. With this, we refer to the use of technologies that

help the production team in performing tasks like structuring content, temporally segmenting content, storing metadata, referencing supplemental material and linkage to associated user-generated content. Content enrichment can be implemented either on existing, non-interactive radio content (i.e., as an afterthought) or could be upfront factored into the radio production processes. A good example of the former approach is given by Cowlishaw et al., who exploit manually entered metadata (e.g., synopses of radio programs) to facilitate the transformation from non-interactive to interactive radio content [15].

Content segmentation Studies for both TV and radio content have shown that users tend to consume program segments rather than entire programs when accessing content on demand. For example, the authors of [37] report that 48% of the respondents in their study found it important to easily locate a part (or section) of a radio program. This means that content needs to be structured into semantically meaningful temporal units. Unfortunately, basic segmentation of audio content typically means segmenting into speech, music and other acoustic elements. By drilling down the resulting disassembled elements (i.e., further analyzing them using topic detection algorithms or by leveraging available EPG data), topical units can be identified and interaction options can be added. Speech is particularly relevant, as interaction opportunities can be added based on the spoken content. For music, interaction may mean adding/linking background information related to the music being played.

Speech to text (ASR) Automatic speech recognition (ASR) is a base technology for other natural language processing tasks, such as named entity recognition, topic detection and clustering. However, the intended purpose of the resulting transcript determines its quality requirements. For further processing, it is important that the semantically relevant terms are correctly captured, while a grammatically completely correct transcription is not needed. If on the other hand the transcript shall also be presented to the user, its quality requirements are significantly higher. The interactive podcast player introduced in [6] visualizes the transcript, and 25% of the users stated it as their preferred feature. Transcript requirements may also depend on the user context, e.g., a full transcript is likely to be useful for users with hearing impairments or who want to exploit the radio content to improve their proficiency in a foreign language.

Topic detection/clustering In cases where detailed structural information is not available from the production process, the topic structure of the content must be recreated. Topic detection and clustering based on the ASR transcript can be applied, and the result can be improved by exploiting external information (e.g., a list of potential topics). Such external information may be helpful even if it is just global, and not associated with the media timeline. If a chapter structure is not only used as a navigation aid for consuming the program but also for consuming content segments independently (i.e., outside the context of the program), then editors need to take care that such segments are self contained. Automated natural language processing tools can support this step by indicating when the main topic of the segment is introduced, or when named entities related to the topic are first mentioned in the entire broadcast [12].

Summarization Summarization aims at creating shorter or condensed versions of content (while maximally retaining the core "message" of the content) in order to facilitate faster content consumption or to yield quick content overviews for browsing purposes. In the radio production context, *dynamic* (or *elastic*) summaries are mostly of interest, in order to enable consumers to access a version of the content that fits both their interest and time constraints. A large body of work on text and video summarization exists, while most work on audio summarization is geared towards music [5] or audio events [44]. Work on summarising spoken content [42] does not aim to create the summary as an audio segment, but only as a text. Hardly any work addresses the problem of generating re-edited audio content, which means posing the problem as one to be jointly solved on audio and text modalities [43]. In addition, scalable summarization technologies, as proposed for video content [27], would be needed to allow consumers to flexibly adjust the desired content length.

Qualitative and affective content attributes Audio signals that are of high technical fidelity (i.e., according to objective metrics like sampling rate or encoding bitrate) are not necessarily positively assessed by listeners from a qualitative point of view. This observation makes the case for enriching radio content with subjective attributes that can have either a positive or negative valence. Examples of interesting qualitative content attributes in the context of speech clips are intonation variation from the speaker (or lack thereof) and the presence of disturbing background noise. Detailed analysis of the audio signal (i.e., inspecting the waveform over time to estimate intonation variation) and machine learning models (as proposed in [21] for user-generated video) could support radio producers by automatically flagging content that exhibits undesirable qualitative attributes. The annotation and exploitation of content attributes however does not need to be limited to quality-related aspects. Again considering the case study of speech clips, other potential approaches to uncover interesting content attributes include sentiment analysis, gender classification of the speaker [25, 28], speaking mode classification (e.g., yelling versus whispering), and auditory emotion derivation (e.g., sad versus happy). Such non-qualitative subjective attributes could be meaningfully applied in the context of, for example, offensive content identification and to match radio content to the mood of the listener (see also Section 3).

Hyperlinking external content A survey [38] found that the most popular radio content genres among Generation C listeners are music, arts & culture, events, and local information. All these genres can benefit from linking the radio content to other information on the web, on social media or with other multimedia content. Especially for younger consumer demographics, the notion of hyperlinks in any content is very natural, and tying in links to related content may avoid that they use other services in parallel to retrieve related information.

One obvious target for linking are sources that were used during the creation of the radio content, such as original articles or interviews that are cited. Ideally, these materials are still available from the production process, otherwise they have to be discovered again (e.g., via web search). The study in [6], dealing with information/documentary programs, found that consumers particularly liked visualizations such as diagrams. Such diagrams can be treated like other external content, i.e. as a graphic file or image.

However, if a data-driven approach to journalism is used, the data analysis and presentation can also be supported by automatic tools [32]. In such a case visualizations of the data can be created on the fly, and can also be adapted to the user context (e.g., screen real estate, desired level of detail, highlighting local data relevant to the consumer).

Hyperlinking of video content is a well studied problem, among others through a series of benchmarks in the MediaEval and TRECVID evaluation campaigns [18]. The aim is to find cross-links in video collections to enable the user to view contextually relevant segments. This approach could of course also be applied across collections and information sources, including the web. Although some approaches strongly rely on audio content (e.g., [23]), there is little work specifically on hyperlinking of audio content [24].

A final potential application of hyperlinking concerns the discovery of social media content relevant to the radio program at hand. Again, this problem has been studied in more detail for linking video, e.g. with social media content related to events [31].

Annotating and filtering user-generated content Wardle and Williams provide a classification of different types of UGC in a broadcasting context, ranging from audio comments to networked journalism, but focus mostly on the use of UGC in relation to news coverage [40]. The contribution of non-news UGC in the context of a radio-supported charity event running for 14 weeks is discussed in [14]. Around 68% of the 24,615 received messages contained multimedia content. This poses challenges of selecting content based on content value, technical quality and diversity. A workflow for content selection is described in [3], using automatic quality analysis and similarity matching tools. With the recent advances in object detection and semantic scene segmentation, there is potential to create more fine-grained and directly usable semantic metadata.

A basic way of presenting related UGC in the radio production workflow has been implemented in [7] as part of the radio coverage of live events such as music festivals. Here, the content is provided as a ranked list, from which the editor can choose elements for inclusion in the radio show. In this example, the rank is learned from past editorial decisions, and the use of online random forests enables adaptation to changing trends [4]. While applying UGC in radio contexts is not widely explored yet, examples of leveraging UGC in video or TV contexts are rather plentiful (e.g., [21, 41]).

3 Context Derivation and Adaptation

Context adaptation involves technologies that can help in better aligning content to the consumption context, either by simplifying the production of multiple, specialized content versions, or by (semi-)automatic content transformation during delivery. Creating new instantiations of existing content or adapting content to the listener's context hinges on tools that are capable of extracting content metadata, information about the user and contextual data in general, and consumption behavior (e.g., audience measurements).

Technical capabilities of target device An important piece of the contextual puzzle is given by the listener's playback device and its technical capabilities. As a simple example, it makes little sense to deliver pre-rendered surround sound to a listener who is

using a pair of stereo headphones in conjunction with a smartphone. Another principal aspect of the consumption device are its display capabilities. If such capabilities are absent, visual supplemental information need not be delivered to the listener (to avoid wasting network bandwidth on information that the user will not be able to consume anyway). On the other hand, if a display is present, its form factor and hardware characteristics are also of interest. Indeed, the physical size and technical capabilities of the display influence the type of visual content that can be meaningfully consumed on it (e.g., still images, long pieces of text like transcripts, geographic maps, charts [6], video). The object-based media paradigm holds the promise to dynamically tune radio content to device capabilities, yet this paradigm is currently still hampered by the challenge that its production processes differ radically from traditional radio production [20].

Environmental context Another invaluable source of contextual information is given by the environment in which the user resides while listening to radio content. Listeners are likely to consume radio differently in solitary versus group settings. When listening together with others, users might be more willing to *go with the flow*, whereas in solitary settings these same users might be more eager to actively engage in content selection. Whether similar behaviors also hold for the witnessed level of interaction with radio content remains to be determined; it might very well be the case that group settings spur radio interactions that would not occur in solitary consumption contexts. Another crucial type of environmental context is the activity the consumer is currently involved in (e.g., working at a desk versus doing a physical workout), and whether or not she is willing or even able to interact. In situations where the user does not have her hands free to perform interaction (e.g., while cooking [16]), smart speaker technology could be helpful [39]. However, smart speakers involve a different way of interacting with radio content as opposed to a typical radio player, which must be taken into account during radio production. The mood of the listener is also likely to affect listener behaviour and is hence valuable contextual information that is best taken into account when delivering radio content [11]. Finally, whether the listener is consuming the radio content live or in an on-demand fashion is also an important type of environmental context. Catching up on radio items has been found to resemble the experience of consuming a podcast, which is a more intimate way of listening compared to live radio [30]. This observation implies that simply hosting (segmented) broadcast content online for on-demand consumption is unlikely to yield optimal listening experiences. Stated differently, live and on-demand radio content demand different, specialized radio production approaches.

Wearable technology, and sensors in general, bring a whole new dimension in gathering analytics about listeners and their consumption environment. A simple example would be to automatically adapt a musical playlist to the current heart rate of a runner. As another example, a seamless handover from IP to DAB+ could automatically be triggered when a user enters her car [17]. However, inferring contextual information that is as intricate as exemplified in the previous paragraph is no small feat and will likely require intelligent reasoning on the fusion of different contextual sources. Also, even if environmental context can be precisely and comprehensively derived, properly adapting radio content to it is far from a solved problem.

Personalization An important form of context-driven content adaptation is personalization. In effect, personalization and recommendation engines take contextual information as input to deliver relevant pieces of content to the user. Content adaptation through personalization can take many forms in a radio consumption context. A straightforward personalization example is to exploit the listener's preferences and profile information to curate a tailored stream of radio items (e.g., a personalized playlist of relevant recent news items as they have been discussed on radio [35, 33]). A more advanced approach would be to replace specific songs in the radio broadcast (using hybrid radio technologies [2]) in accordance with the musical taste of the listener. Personalization can also take place on different levels of individualization. In the most extreme case, personalization is intended to serve a single listener. However, personalization could also cater to groups of like-minded listeners, or even to the radio community as a whole. A real-world example of community-driven personalization is given by the concept of listener curation of the radio playlist, by letting listener voting determine the songs that will be played [29].

Interaction can play an important role in personalized radio experiences. As an example, listeners could provide feedback about the quality and relevance of the recommendations they receive [15]. In certain personalization schemes, interaction is even indispensible, as is exemplified by the use case of listener curation of the radio playlist (see previous paragraph).

A common challenge with personalization and recommendation is the filter bubble effect. By exclusively catering to the taste and preferences of the listener, the risk arises that the listener gets "locked in" in his or her personal interest sphere or, more worrying, that the listener's view on the world becomes too one-sided. This is of particular interest in the context of public service radio, given that fostering diversity and pluralism is one of the six core values of public service media [19]. Personalization and recommendation are furthermore known to have an adverse effect on serendipity, which is still claimed to be one of broadcast radio's strong suits [15]. Finally, by design, radio is intended to deliver a shared, uniting experience to entire communities; clearly, personalization undermines this objective. Finding the sweet spot between serving the community as a whole and catering to the increasing personalization desires from listeners is an important open challenge in radio research.

Adjusting content duration Consumers often have a limited time slot to fill when they turn to the catch-up consumption of radio content (e.g., while commuting). A specialized form of radio personalization therefore revolves around content duration adaptation [33]. The goal of such adaptation approaches is to tell a story that is as complete and as compelling as possible, within a certain time limit. Given that it is practically infeasible to prepare different versions of the story for each possible time budget, coming up with algorithmic solutions that on-the-fly populate the available time slot in an optimal fashion is an important area of future research (cf. the Squeezebox prototype described in [20]).

Deriving context from listener input A method for users to interact with radio stations that became popular recently is through chatbots that serve as a first line of support [14]. The resulting conversations can lead to novel ways of profiling and

analyzing listeners. Consider a listener that queries a chatbot about the currently playing song. After responding with the requested song title and artist information, the chatbot automatically follows up with the question whether the requesting listener actually likes the song. Based on the listener's response to this question (if any), his or her profile information can be refined (specifically with respect to musical preferences), this way allowing the broadcaster to gather deeper knowledge about its listeners. Another important form of listener input in the context of radio interaction is UGC (see Section 2). Automatically scanning and analysing such UGC can also provide clues about the user's current context (e.g., location, activity, mood).

4 Production User Interfaces

Radio production user interfaces need to be updated to integrate the rich information provided by automatic tools and present it in a way that enables efficient production processes.

Live versus nearly live consumption data When radio is consumed through a website or application, radio producers are already enabled to monitor listeners' experiences, using off-the-shelf tools as Adobe Analytics[7] or tailored solutions as Voizzup[8]. The latter experimented with offering listener consumption data in real-time to the radio production team, yet learned that radio hosts felt it affected the production of their show. For instance, learning that a subset of listeners is disconnecting during a particular interview does not encourage a radio host to continue that interview, although it still might be relevant. As a result, the tool is developed to disclose consumption data only right after the end of the radio show. Whether this approach of nearly live data delivery also applies to other types of data, such as contextual information (see Section 3), is still an open question.

Visualization and transparency In a live production environment, data has to be interpreted fast. Text messages (e.g., contributed by listeners through chat or SMS), for instance, provide a rich source of stories that could be told on radio. Nowadays, such stories commonly have to be spotted and distilled manually by radio producers. When text messages contain photos, it becomes easier for the radio producer to interpret and assess the potential story in a fast way [13]. Because of such practical considerations, in present-day radio production, visually augmented listener messages have a higher probability of being included in the radio broadcast. This in turn indicates the potential of enriching data with automatic visual annotations (e.g., a map of geo-tagged text or photos submitted by listeners). Furthermore, as novel types of data (such as those mentioned in previous sections) are being tracked and cause automatic content adaptations, there is a need to present those decisions in an easy and fast-to-understand format to radio producers. Overall, intelligent and transparent data visualization in radio production interfaces should allow radio producers to stay in control to ensure serendipity (as mentioned in Section 3) and storytelling quality.

[7] `https://www.adobe.com/analytics/adobe-analytics.html`
[8] `http://www.voizzup.com/`

Efficient overview Radio producers are currently confronted with a multitude of tools, software, interfaces, and so on during the live production process. As a result, they are often distracted from their core content production tasks (e.g., when searching for the right tool to contact a listener) [8]. In future, a smart production system might automatically reconfigure its user interface by visualizing only the data, content and context that is relevant in the current phase of the radio production process or given the current radio production task at hand [22]. Conversely, currently irrelevant information or tools could be hidden, but remain accessible to the production team when needed. Automatic interface reconfiguration decisions could be inspired by, for example, the pre-produced script of the radio show.

5 Conclusion

In this position paper, we have analyzed challenges for producing interactive radio content, which fits the need of contemporary *any where, any time, any device* media consumption habits. On the one hand, the production team needs to stay in control to create high quality radio experiences. On the other hand, automation support is needed to handle the large amounts of heterogeneous multimedia content that is involved in interactive radio production. We argue that, to some extent, interactive radio production is in fact a multimedia analytics problem. We have analyzed three major aspects in interactive radio production (i.e., content enrichment, context derivation and adaptation, and the integration in production user interfaces) and have uncovered important challenges and opportunities concerning the use of multimedia analytics for each of them. In this process, we have also identified important research gaps in the analysis and processing of radio content. Equivalent research has commonly already been conducted for media types other than radio (mostly video); it is worthwhile to investigate whether these findings from other media domains are translatable to interactive radio contexts.

Acknowledgments. This work has received funding from the European Union's Horizon 2020 research and innovation programme, under grant agreement n° 761802 MARCONI ("Multimedia and Augmented Radio Creation: Online, iNteractive, Individual"). The authors thank all presenters and participants of the Workshop on Interactive Radio Experiences at ACM TVX 2019.

References

1. Media use in the European Union. Tech. rep., Standard Eurobarometer 88 - Wave EB88.3 (Nov, 2017)
2. D2.1: HRADIO User Scenarios. Tech. rep., H2020 HRADIO Project Deliverable (2018)
3. Bailer, W., Thaler, M., Horti, A., Grandl, R., Hofmeister, W., Steiner, J., Fink, H.: Content and Metadata Workflow for User Generated Content in Live Production. In: Proc. European Conference on Visual Media Production. London, UK (2016)
4. Bailer, W., Winter, M., Wechtitsch, S.: Learning selection of user generated event videos. In: Proc. 15th International Workshop on Content-Based Multimedia Indexing (2017)

5. Bartsch, M.A., Wakefield, G.H.: Audio thumbnailing of popular music using chroma-based representations. IEEE Transactions on multimedia **7**(1) (2005)
6. Baume, C.: "Even More or Less": A data-rich interactive podcast player. In: Proc. Workshop on Interactive Radio Experiences at ACM TVX (2019)
7. Bauwens, R.: Interactive Content Contribution. In: Proc. Workshop on Interactive Content Consumption at ACM TVX (2016)
8. Bauwens, R., Claes, S.: MARCONI: Towards an integrated, intelligent radio production platform. In: Proc. IBC2019 Conference (to appear) (2019)
9. Bauwens, R., Lievens, H., Wijnants, M., Pike, C., Jennes, I., Bailer, W.: Interactive Radio Experiences. In: Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video (TVX) (2019)
10. Bois, R., Gravier, G., Jamet, E., Morin, E., Robert, M., Sébillot, P.: Linking multimedia content for efficient news browsing. In: ACM Intl. Conf. Multimedia Retrieval (2017)
11. Buurman, A.: Audio draait om mood management. `https://beautifullives.com/news/audio-draait-om-mood-management/` (2016)
12. Chen, Y., Liu, L.: Development and research of topic detection and tracking. In: Proc. IEEE International Conference on Software Engineering and Service Science (2016)
13. Claes, S., Bauwens, R., Matton, M.: Augmenting the Radio Experience by Enhancing Interactions Between Radio Editors and Listeners. In: Proc. ACM International Conference on Interactive Experiences for TV and Online Video (TVX) (2018)
14. Claes, S., Jansen, S.: 360 Radio Experience: Connecting Live Listening with User Interaction and Visual Radio. In: Proc. Workshop on Interactive Radio Experiences at ACM TVX (2019)
15. Cowlishaw, T., Burlington, T., Man, D., Fiala, J., Barrington, R., Wright, G.: Personalising the public: Personalising linear radio at a public service broadcaster. In: Proceedings of the IBC2018 Conference (2018)
16. Cox, J., Jones, R., Northwood, C., Tutcher, J., Robinson, B.: Object-based production: A personalised interactive cooking application. In: Adjunct Proc. ACM International Conference on Interactive Experiences for TV and Online Video (TVX) (2017)
17. Erk, A., Sattler, F.: DAB over IP. In: Proc. Workshop on Interactive Radio Experiences at ACM TVX (2019)
18. Eskevich, M., Jones, G.J., Aly, R., Ordelman, R.J., Chen, S., Nadeem, D., Guinaudeau, C., Gravier, G., Sébillot, P., De Nies, T., et al.: Multimedia information seeking through search and hyperlinking. In: Proceedings of the 3rd ACM International Conference on Multimedia Retrieval (2013)
19. European Broadcasting Union (EBU): Empowering Society - A Declaration of the Core Values of Public Service Media. `https://www.ebu.ch/files/live/sites/ebu/files/Publications/EBU-Empowering-Society_EN.pdf` (2012)
20. Evans, M., Ferne, T., Watson, Z., Melchior, F., Brooks, M., Stenton, P., Forrester, I., Baume, C.: Creating Object-Based Experiences in the Real World. SMPTE Motion Imaging Journal **126**(6) (2017)
21. Evans, M., Kerlin, L., Larner, O., Campbell, R.: Feels Like Being There: Viewers Describe the Quality of Experience of Festival Video Using Their Own Words. In: Extended Abstracts of the ACM Conference on Human Factors in Computing Systems (CHI) (2018)
22. Gajos, K., Weld, D.S.: SUPPLE: Automatically Generating User Interfaces. In: Proceedings of the 9th International Conference on Intelligent User Interfaces. pp. 93–100. IUI '04, ACM (2004)
23. Galuščáková, P., Pecina, P.: Audio Information for Hyperlinking of TV Content. In: Proc. 3rd Workshop on Speech, Language & Audio in Multimedia (2015)
24. Gehani, S.: Audio Hyperlinking. US Patent App. 13/368,129 (2013)

25. Gemmeke, J.F., Ellis, D.P.W., Freedman, D., Jansen, A., Lawrence, W., Moore, R.C., Plakal, M., Ritter, M.: Audio Set: An ontology and human-labeled dataset for audio events. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP) (2017)
26. Gravier, G., Ragot, M., Amsaleg, L., Bois, R., Jadi, G., Jamet, E., Monceaux, L., Sébillot, P.: Shaping-up multimedia analytics: Needs and expectations of media professionals. In: International Conference on Multimedia Modeling (2016)
27. Herranz, L., Martinez, J.M.: A framework for scalable summarization of video. IEEE Transactions on Circuits and Systems for Video Technology **20**(9) (2010)
28. Hershey, S., Chaudhuri, S., Ellis, D.P.W., Gemmeke, J.F., Jansen, A., Moore, C., Plakal, M., Platt, D., Saurous, R.A., Seybold, B., Slaney, M., Weiss, R., Wilson, K.: CNN Architectures for Large-Scale Audio Classification. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP) (2017)
29. Jackson, W.: Listener Driven Radio Gains Popularity. `https://www.radioworld.com/news-and-business/listener-driven-radio-gains-popularity` (2018)
30. Krotoski, A.: Touchies and Feelies: Everything I Know About Human Interfaces. ACM CHI Opening Keynote (2019)
31. Liu, X., Huet, B.: Linking socially contributed media with events. Multimedia Systems **22**(4) (2016)
32. Montagnuolo, M., Messina, A.: Rai Active News: Integrating Knowledge in Newsrooms. In: Proceedings of the IBC2015 Conference (2015)
33. Paolo Casagranda, F.R., Prioletti, R.T.: Personalising linear radio: Model and user evaluation. In: Proc. IBC2018 Conference (2018)
34. RAJAR/IpsosMori: MIDAS Audio Survey Autumn 2018. Tech. rep. (2018)
35. Sappelli, M., Chu, D.M., Cambel, B., Nortier, J., Graus, D.: SMART Radio: Personalized News Radio. In: Proc. 17th Dutch-Belgian Information Retrieval workshop (DIR) (2018)
36. Silzle, A., Weitnauer, M., Warusfel, O., Bleisteiner, W., Herberger, T., Epain, N., Duval, B., Bogaards, N., Baume, C., Herzog, U., et al.: ORPHEUS Audio Project: Piloting an End-to-End Object-based Audio Broadcasting Chain. In: Proc. IBC2017 Conference (2017)
37. Skov, M., Lykke, M.: Unlocking radio broadcasts: user needs in sound retrieval. In: Proceedings of the 4th Information Interaction in Context Symposium (2012)
38. Starkey, G., Gazi, A., Dimitrakopoulou, D., Cordeiro, P.: "Generation C" and audio media: A cross-cultural analysis of differences in media consumption in four European countries. Participations: Journal of Audience and Reception Studies **11**(2) (2014)
39. VRT Innovation: "Alexa, what's the news ?" VRT is experimenting with voice-controlled radio and news. `https://innovatie.vrt.be/en/en/article/vrt-is-experimenting-with-voice-controlled-radio-and-news` (2018)
40. Wardle, C., Williams, A.: Beyond user-generated content: A production study examining the ways in which UGC is used at the BBC. Media, culture & society **32**(5) (2010)
41. Wijnants, M., Rovelo, G., Quax, P., Lamotte, W.: Wandercouch. Multimedia Tools and Applications **76**(4) (2017)
42. Zechner, K.: Automatic summarization of open-domain multiparty dialogues in diverse genres. Computational Linguistics **28**(4) (2002)
43. Zlatintsi, A., Iosif, E., Marago, P., Potamianos, A.: Audio salient event detection and summarization using audio and text modalities. In: Proceedings of the 23rd European Signal Processing Conference (EUSIPCO) (2015)
44. Zlatintsi, A., Maragos, P., Potamianos, A., Evangelopoulos, G.: A saliency-based approach to audio event detection and summarization. In: Proceedings of the 20th European Signal Processing Conference (EUSIPCO) (2012)