

CROWDSOURCING AUDIENCE PERSPECTIVES ON CLASSICAL MUSIC

Cynthia C. S. Liem

Multimedia Computing Group
Delft University of Technology, The Netherlands
c.c.s.liem@tudelft.nl

ABSTRACT

Program notes of (Western) classical music are usually written by musical experts. As a consequence, program notes information may not be very accessible to audiences which are less familiar with this music genre. To gain more insight in the way in which ‘uninitiated’ audience perceives and describes symphonic concert performances, in this work, we describe the acquisition of a dataset of various forms of timeline commentary through crowdsourcing mechanisms. As we show, different audience categories use different vocabularies, and consider different anchors on the timeline to be important.

Index Terms— music, concerts, commentary, tagging, crowdsourcing

1. INTRODUCTION

The genre of Western classical music forms an important part of the cultural heritage of Europe. However, in recent times of insecure funding and extensive competing online offer of all kinds of music, policy makers, music institutions and musicians share concern on the sustainability of their audiences.

In recent history, classical music may have grown to have an ‘elitist’ image, with the core audience typically being older and highly educated [1]. The rituals and common vocabulary used around the genre may not be known to ‘outsider’ audiences with potential interest in the music. This is a risk, as a feeling of lack of knowledge of the music and the etiquette surrounding it will hamper the experience of a concert [2, 3].

In reaction to this development, several digital solutions have recently been proposed to enrich and elucidate what is happening in the music. For example, the LiveNote Orchestral Performance Companion [4] as well as the Wolfgang app¹ aim to assist audience in a live concert performance by giving relevant synchronized background information on what is happening in the music.

Still, it is not trivial that simply enabling richer digital servicing will immediately lead to improved user appreciation, understanding and inclusion. For this, it also is important that

digital services will make use of appropriate content, and that this content is served in the right way.

As for synchronized background information applications, the provided information forms an alternative to paper program notes, and usually is written by music experts. This means that this information may not necessarily be in the natural vocabulary of audiences who are less familiar with the classical music genre.

In the current paper, we aim to gain novel insights into the way in which non-insider audiences to classical music actually describe music of the genre. To this end, we describe the crowdsourced acquisition and initial analysis of a dataset [5]², offering free-text commentary by users at various expertise levels, in reaction to an audiovisual recording of Beethoven’s Eroica symphony. We describe how the crowdsourcing task was set up, and what the main characteristics of the resulting user vocabulary are. Besides, we also provide an initial investigation of *anchoring* effects: the localization on the concert recording timeline of events which were experienced as particularly striking by a user.

The remainder of the paper is organized as follows. In Section 2, we first discuss the research context of our work, also offering a broader perspective on relevant related work in Section 3. After this, in Section 4 the crowdsourcing task setup for acquiring user input is explained, followed by an overview of task completion statistics in Section 5. Main findings from an initial data analysis are discussed in Section 6, after which we will provide final conclusions.

2. CONTEXT: THE PHENICX PROJECT

The currently presented work resulted from the recently concluded European research project PHENICX [6]. In line with the various efforts to innovate digital concert services as mentioned in our Introduction, PHENICX aimed at using digital technology to multimodally enrich the experience of classical concert performances, before, during and after the concert event takes place. This led both to algorithmic advances in the extraction of relevant musical information from rich and

¹<http://wolfgangapp.nl/en/>

²publicly available at <http://doi.org/10.4121/uuid:f1b505d9-71cc-4a68-835b-3b6fbffc131b>

multimodal information streams [7], but also to the establishment of several use cases which were of interest from an academic, industrial and musician perspective [8]. Besides, user-centered design and development was extensively applied [3] such that prototypical outcomes integrating novel technologies would indeed match requirements of users and be effective in raising enthusiasm about classical music.

An important goal of PHENICX was to broaden the audience for classical music. As a possible mechanism to do this, social sharing mechanisms were foreseen among the use cases ('sharing the magic') [8]. Here, the idea would be that existing concertgoers could share favorite parts of concerts with their friends, and post comments about the music on the concert timeline. This functionality was implemented in a tablet prototype [9] that could combine timed comments of regular concertgoers with that of experts, resulting in customized information feeds.

However, we found that this functionality was somewhat problematic: entering extensive comments proved too much of effort for user audiences in live settings, and audiences did generally not show immediate willingness to produce and share. In general, possibly due to classical music audiences typically being older, social media mechanisms and presence are not as widespread for this genre as in other genres [10]. Besides, attracting and interesting outsider audiences was in general harder than doing this for so-called heavy and casual consumers: audience members who have some (up to extensive) familiarity with classical music already. User experience evaluations of the PHENICX prototypes indicated a knowledge gap, with outsider audiences not always understanding the vocabulary used in the music commentary.

As an alternative, in [11] we proposed a personalized tag recommendation system, considering tags produced by the crowd, and personalizing based on personality-related factors. An advantage of using tags is that they are entered quickly; however, a disadvantage is that they are not very information-rich, and tend to describe 'obvious' concepts.

These observations led to the current work. On one hand, it seeks richer descriptions over the timeline of the concert, from more varied user audiences at various expertise levels. On the other hand, we do not aim to perform a user experience study at this stage. Instead, we choose to focus on an analysis of the formulation of descriptions based on a recording, but not on interaction input mechanisms on a mobile device during a live concert.

3. OTHER RELATED WORK

Our work can be related to efforts broader than the context of digital concert services. The search for vocabularies supported by the crowd considers folksonomy analysis. As an example of such analysis in the music domain, in [12], it was shown that consensus of expert vocabularies and community-contributed folksonomies from social music services was not

always found for the description of musical genres.

Typically, folksonomies are established in social tagging contexts. The acquisition of decently sized corpora of such tags (outside of music services, such as `last.fm`) was typically performed through games with a purpose mechanisms, e.g. in TagATune [13], MajorMiner [14] and the ListenGame [15]. In such games, consensus with existing user contributions was used as part of the scoring mechanisms, favoring consistently occurring, broadly supported taggings. An important assumption underlying the game scoring mechanisms is that tags which are mentioned by multiple users are the most reliable.

However, such approaches to folksonomic description acquisition have two major restrictions, which are undesired in the context of synchronized concert information. First of all, tags are restricted to very short keyword-like descriptions, which would not read like a natural sentence. Secondly, if consensus is used as quality criterion, obvious but general tags will stand out which may not be as interesting as complementary information to a concert experience.

In earlier work focusing on narrative associations to production music [16], we experimented with crowdsourced free-text descriptions, primed from a creative writing assignment, as a richer alternative to folksonomic tags. The methodology in the current work is partially inspired by this work.

Another important interest of ours is the localization of striking events on a concert timeline, which would ultimately enable non-linear access mechanisms within a longer digital recording. This relates to recent emerging interests in the Multimedia community on video search and hyperlinking and automatic anchor inference for navigation across different documents in archives [17, 18, 19]. Referring back to the project use case targeted at social sharing and discussing of interesting moments in a concert, those parts on the timeline that will be suitable for this purpose can be considered as anchors too.

4. TASK SETUP

As mentioned in the previous section, following good prior experiences with larger crowdsourcing assignments (HITs) on Amazon Mechanical Turk involving free-text descriptions in previous work [16], we designed a HIT which also would solicit musical descriptions from users in a free-form way.

We used an audiovisual recording of a performance of Beethoven's Third Symphony 'Eroica', performed by the Royal Concertgebouw Orchestra conducted by Iván Fischer in May 2013. As we planned to obtain full time line coverage of the recording (in total around 55 minutes), we split the video up into segments of 2 minutes. Crowdworkers were asked to watch the full 2-minute video. Subsequently, they were asked to give a description of the video fragment which *a friend with similar musical expertise to themselves* would understand.

Then, crowdworkers were invited to indicate and characterize *particularly striking points* on the fragment time line. After that, they were asked to indicate points that *stood out to them, but which they would think others would not obviously notice*. This last question was added to encourage crowdworkers to not only look for obvious, high-consensus input, but also to other elements that they personally found significant.

The HIT started with an informed consent notice and a short background questionnaire. After the main assignments listed above, a short feedback questionnaire was provided in which workers could indicate how difficult they found each of the assignments. The HIT was first piloted and then released to acquire 10 distinct responses per fragment, at \$0.75 reward per assignment.

5. TASK STATISTICS

The crowdsourcing task completed within 48 hours. 271 HIT assignments were accepted (amounting to 233 unique workers: 117 female and 112 male), while 30 HIT assignments were rejected (amounting to 12 unique workers). Not all of the accepted HIT assignments were fully completed; we still accepted submitted assignments in case workers indicated any particular issue in completing the assignment (such as technical problems). In the section on our initial data findings, we will further elaborate on the amount of data obtained for each sub-assignment of the HIT.

Age category	count
under 20	3
20-29	87
30-39	76
40-49	28
50-59	27
60-69	9
70 or older	2
unknown	1

Table 1: Age distribution of workers

Analyzing the pool of accepted work, as can be seen from Table 1, the 233 workers were relatively young. Most workers (207) spent the formative years of their life in the USA, 17 in India, 9 in other countries. A histogram of HIT completion times is given in Figure 1.

As indicated above, we distinguish between three audience groups of interest: outsiders, casual consumers and heavy consumers. To assign these categories to the workers, as part of the background questionnaire, we had participants picking the best-fitting description of their relationship to classical music:

- “This is not the music I’m mainly interested in. I would normally not attend classical concerts.” *People picking*

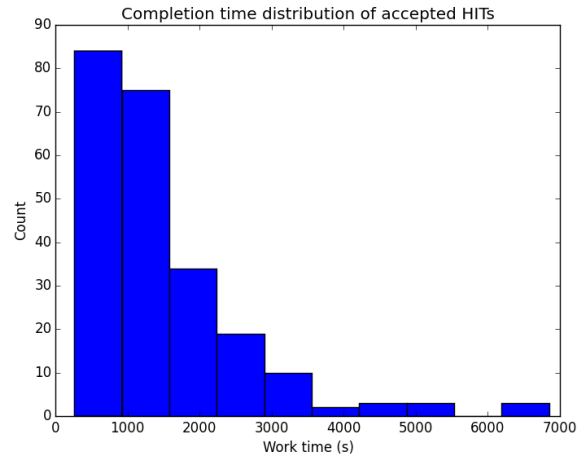


Fig. 1: Histogram of HIT completion times.

this description were labeled as ‘outsider’.

- “I’m quite ok with classical music. Sometimes I also attend classical concerts, but I do not consider myself to have deep classical music expertise.” *People picking this description were labeled as ‘casual consumer’.*
- “I’m a real classical music aficionado! I frequently attend classical concerts and consider myself an expert when talking about the genre.” *People picking this description were labeled as ‘heavy consumer’.*

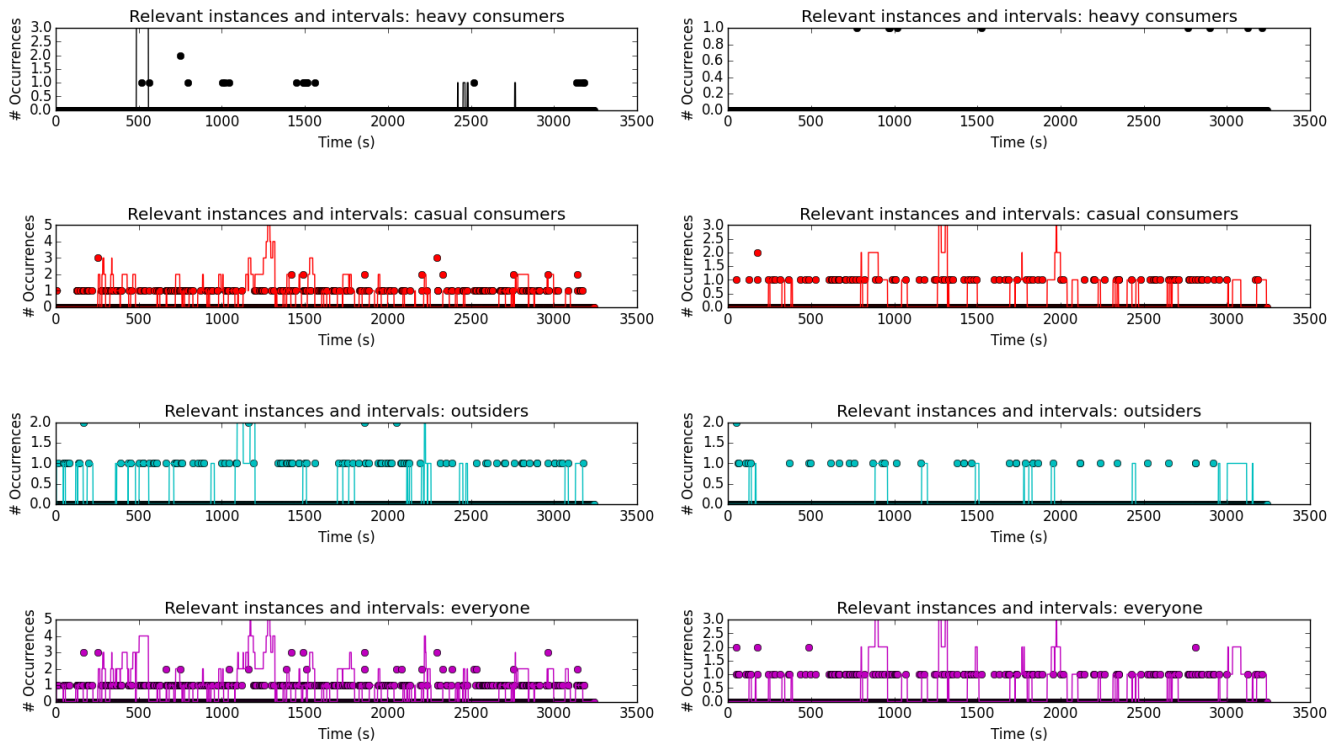
From the responses to these descriptions, 16 workers reported as heavy consumers, 148 workers as casual consumers, 67 as outsiders, and 2 declined to answer. In the remainder of this paper, we will particularly focus on responses given by casual and outsider consumers.

6. MAIN FINDINGS

In this section, we will present the main findings of a first analysis of our dataset, investigating vocabulary and topics of interest regarding the descriptions of the various fragments of the symphony; anchoring behavior regarding the indication of striking points; and finally, anchoring behavior regarding the indication of ‘non-obvious’ points. For each of these findings, we will provide a short outlook on how they can be used for further research.

6.1. Overall time line descriptions

In total, we received 263 free-text descriptions of video fragments. Workers had relatively little trouble in producing these descriptions: on a scale from -2 (very difficult) to 2 (very easy), the average difficulty rating for this assignment was -0.16.



(a) 'Points standing out'

(b) 'Non-obvious points standing out'

Fig. 2: Instances and intervals of relevance for individual and aggregated audience types.

From a qualitative inspection of the provided descriptions we gained several insights:

- Workers frequently describe fragments in terms of tempo (fast, slow) and dynamics (loud, soft) (“*It started somewhat soft and slow. It gradually grew somewhat faster and louder. Then it got really loud towards the end.*”). Another frequently mentioned term is that of *intensity*, which may relate to these concepts in a more abstract sense (“*The intensity really builds and things really take a kind of dark tone overall.*”).
- Workers do not often use technical vocabulary for structural elements (e.g. themes), or may use mistaken vocabulary. At the same time, they use figurative language to describe developments (“*The key of the fast melodies often started as low, then went to a high point, then it descended to a low point again as if it was drawing a line shaped like a mountain.*”, “*This piece ebbs and flows.*”).
- In contrast, music-technical terms that occur frequently include ‘staccato’ and ‘crescendo’. (“*The segment starts with a staccato build for a few seconds. It then*

dances between slow and fast segments of varying intensity. It ends with a loud and rapid build with what feels like a pursuit.”).

- Several times, workers mention that the music reminds them of film music (“*The music seemed to tell a story, reaching a height and then falling. It very much resembled a score of a film in that sense.*”).
- Following up on the previous point, workers may describe their impressions in the form of a non-musical story (“*I imagine a dog lost in the streets of a bustling city. It’s raining and he is kicked and shoved out of the way of many passersby in raincoats and big splashing boots. He’s scared.*”).
- The coordination between orchestra players stands out and is frequently described, both in visual (“*The musicians all played their instruments with desire and coordination with one goal in mind.*”) as well as in musical terms (“*The lower, bass instruments would play an open phrase, and the higher pitched instruments would “respond”.*”, “*From the beginning it seems that the string section is pondering something which is like*

a question building in intensity that is finally asked by the brass section. Then the woodwinds attempt to answer the question. It seems as if the bass section repeats the answer but isn't confident and so the woodwinds continue more adamantly while the strings continue to push the issue until the brass and percussion join in the answer.") The quality of the musicians also is praised multiple times. Interestingly enough, these performer observation-related description aspects closely match several intrinsic motivators for classical concert attendance mentioned from earlier studies [3].

- Workers include own emotional reactions in their descriptions ("Emotionally it made me feel hopeful or optimistic.").
- Workers do not always know that a conductor is called 'conductor' (other used terms include "host", "leader" and even "composer"), but clearly identify the conductor as the coordinating person in the ensemble.

It will be interesting to investigate the findings above in relation to signal properties. Several of the mentioned aspects can be extracted from audiovisual signals, while the more abstract notions of development may need different feature models beyond current standard music-technical descriptors.

6.2. Indicating striking points

Indicating time markers of striking moments was found easiest of all subtasks of the HIT, with an average difficulty rating of 0.07. In total, we collected 544 indicated time markers of striking points from workers over the video fragments. Out of these, 367 markers were indicated to consider a time instance, while 177 markers actually were indicated to consider a longer fragment.

In Figure 2a, we visualized the time markers indicated per audience category. Time locations of important instances are indicated as points. As for fragments, we plot information from fragments for which workers indicated a start time and ending time: if t is a vector representing the timeline of the piece in seconds (from 1 until the length of the piece), $f(t_s, t_e)$ is an indicated relevant fragment starting at t_s and ending at t_e , and $w(t)$ is an occurrence frequency function over the timeline of the piece, each fragment adds a weight of 1 to $w(t_s), w(t_{s+1}), \dots, w(t_e)$.

A first inspection of the data reveals that different types of consumers display differing agreement on relevant fragments and intervals. Instances and intervals chosen by more than 2 workers occur at different points in time for the different audience categories.

It should be verified whether this agreement behavior is triggered by signal-specific properties. Besides, it will be useful to perform a deeper analysis on local distributions of indicated time markers, to analyze whether peaks can be found by aggregating information in local temporal neighborhoods.

6.3. Indicating non-obvious points

Indicating time markers of non-obvious striking moments was a less trivial task. This reflected both in the average difficulty rating (-0.25), as well as in the amount of input received for this task: 208 markers in total, of which 145 were indicated as instances and 63 as fragments.

In similar fashion as described in the previous subsection, we visualize time marker information for this task in Figure 2b. While this data is smaller in scale than that of the previous subsection, again, peaks do not co-occur between the different audience groups, again suggesting different relevance criteria. Besides, it appears that responses to this task create slightly different relevant point indicators than those yielded by the previous subtask of just indicating 'striking' points. At the same time, this can be an artifact of the acquisition procedure (workers did not know about this question until after they completed the previous subtask), so further work needs to be performed in assessing to what extent workers would naturally go for mainstream or non-mainstream types of descriptions.

7. CONCLUSIONS

We described the acquisition of a dataset of crowdsourced audience perspectives on classical music, and provided a first analysis of this dataset with focus on topics of relevance in description, and anchoring behavior when considering time points that stood out, also in non-obvious ways. As most of the workers represented younger and less-initiated audience for classical music, we believe that our dataset valuably contributes to better understanding the language and perception of these groups which are not commonly in the concert hall.

In our discussion, we pointed out how current insights can be used for future research linking signal analysis of the audiovisually recorded concert to the descriptions and observations as provided by the workers. Good linking of these information resources will be essential for offering better tailored information services to broader audiences, and finding more comprehensive relevance criteria for digital concert data exploration.

Acknowledgements: The research leading to these results has received funding from the European Union Seventh Framework Programme FP7 / 2007–2013 through the PHENICX project under Grant Agreement no. 601166.

8. REFERENCES

- [1] H. Roose, "Many-Voiced or Unisono?: An Inquiry into Motives for Attendance and Aesthetic Dispositions of the Audience Attending Classical Concerts," *Acta Sociologica*, vol. 51, no. 3, pp. 237–253, 2008.

- [2] M. C. Dobson, “New audiences for classical music: The experiences of non-attenders at live orchestral concerts,” *Journal of New Music Research*, vol. 39, no. 2, pp. 111–124, 2010.
- [3] M. S. Melenhorst and C. C. S. Liem, “Put the Concert Attendee in the Spotlight. A User-Centered Design and Development approach for Classical Concert Applications.,” in *Proceedings of the 16th Conference of the International Society for Music Information Retrieval (ISMIR)*, Málaga, Spain, 2015, pp. 800–806.
- [4] M. Prockup, D. Grunberg, A. Hrybyk, and Y. E. Kim, “Orchestral Performance Companion: Using Real-Time Audio to Score Alignment,” *IEEE MultiMedia*, vol. 20, pp. 52–60, April–June 2013.
- [5] C. C. S. Liem, “Crowdsourced time commentary on an audiovisual recording of Beethoven’s Eroica symphony,” dataset, publicly available at <http://doi.org/10.4121/uuid:f1b505d9-71cc-4a68-835b-3b6fbffc131b>, 2016.
- [6] C. C. S. Liem, E. Gómez, and M. Schedl, “PHENICX: Innovating the Classical Music Experience,” in *Proceedings of the 2015 IEEE International Conference on Multimedia and Expo (ICME) (EU-Projects Program)*, Torino, Italy, June 2015.
- [7] E. Gómez, M. Grachten, A. Hanjalic, J. Janer, S. Jordà, C. F. Julià, C. C. S. Liem, A. Martorell, M. Schedl, and G. Widmer, “PHENICX: Performances as Highly Enriched and Interactive Concert Experiences,” in *Proceedings of the SMAC Stockholm Music Acoustics Conference 2013 and SMC Sound and Music Computing Conference 2013*, Stockholm, Sweden, August 2013.
- [8] C. C. S. Liem, R. van der Sterren, M. van Tilburg, A. Sarasúa, J. J. Bosch, J. Janer, M. S. Melenhorst, E. Gómez, and A. Hanjalic, “Innovating the Classical Music Experience in the PHENICX Project: Use Cases and Initial User Feedback,” in *Proceedings of the 1st International Workshop on Interactive Content Consumption (WSICC) at EuroITV 2013*, Como, Italy, June 2013.
- [9] M. S. Melenhorst, R. van der Sterren, A. Arzt, A. Martorell, and C. C. S. Liem, “A Tablet App to Enrich the Live and Post-Live Experience of Classical Concerts,” in *Proceedings of the 1st International Workshop on Interactive Content Consumption (WSICC) at TVX 2015*, Brussels, Belgium, June 2015.
- [10] M. Schedl and M. Tkalčić, “Genre-based analysis of social media data on music listening behavior,” in *Proceedings of the 1st ACM International Workshop on Internet-Scale Multimedia Management (ISMM 2014)*, Orlando, FL, USA, November 2014.
- [11] M. Tkalčić, M. Schedl, C. C. S. Liem, and M. S. Melenhorst, “Personalized retrieval and browsing of classical music and supporting multimedia material,” in *Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR)*, New York, USA, June 2016.
- [12] M. Sordo, O. Celma, M. Blech, and E. Gaus, “The Quest for Musical Genres: Do the Experts and the Wisdom of Crowds Agree?,” in *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR)*, Philadelphia, USA, 2008.
- [13] E. L. M. Law, L. Von Ahn, R. B. Dannenberg, and M. Crawford, “TagATune: A Game for Music and Sound Annotation,” in *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, 2007.
- [14] M. I. Mandel and D. P. W. Ellis, “A web-based game for collecting music metadata,” *Journal of New Music Research*, vol. 37, no. 2, pp. 151–165, 2008.
- [15] D. Turnbull, R. Liu, L. Barrington, and G. Lanckriet, “A game-based approach for collecting semantic annotations of music,” in *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, 2007.
- [16] C. C. S. Liem, M. A. Larson, and A. Hanjalic, “When Music Makes a Scene — Characterizing Music in Multimedia Contexts via User Scene Descriptions,” *International Journal of Multimedia Information Retrieval*, vol. 2, pp. 15–30, 2013.
- [17] M. Eskevich, G. J. F. Jones, R. Aly, R. J. F. Ordelman, S. Chen, D. Nadeem, C. Guinaudeau, G. Gravier, P. Sébillot, T. de Nies, P. Debevere, R. Van de Walle, P. Galuscakova, P. Pecina, and M. Larson, “Multimedia information seeking through search and hyperlinking,” in *Proceedings of the 3rd ACM Conference on International Conference on Multimedia Retrieval*, New York, NY, USA, 2013, ICMR ’13, pp. 287–294, ACM.
- [18] R. J. F. Ordelman, M. Eskevich, R. Aly, B. Huet, and G. Jones, “Defining and evaluating video hyperlinking for navigating multimedia archives,” in *Proceedings of the 24th International Conference on World Wide Web*, New York, NY, USA, 2015, WWW ’15 Companion, pp. 727–732, ACM.
- [19] M. Eskevich, R. B. N. Aly, R. J. F. Ordelman, D. Racca, Shu Chen, and G. J. F. Jones, “Sava at mediaeval 2015: Search and anchoring in video archives,” in *Working Notes Proceedings of the MediaEval 2015 Workshop*, Aachen, Germany, 2015, vol. 1436, pp. 11–13, CEUR-WS.org, ISSN=1613-0073.