ETHNOGRAPHIC OBSERVATIONS OF MUSICOLOGISTS AT THE BRITISH LIBRARY: IMPLICATIONS FOR MUSIC INFORMATION RETRIEVAL

Mathieu Barthet
Centre for Digital Music
Queen Mary University of London
mathieu.barthet@eecs.qmul.ac.uk

Simon Dixon
Centre for Digital Music
Queen Mary University of London
simon.dixon@eecs.qmul.ac.uk

ABSTRACT

Without a rich understanding of user behaviours and needs, music information retrieval (MIR) systems might not be ideally suited to their potential users. In this study, we followed an ethnographic methodology to elicit some of the strategies used by musicologists to explore and document musical performances, in order to investigate if and how technologies could enhance such a process. Observations of musicologists studying historical recordings of classical music were conducted at the British Library. The observations show that the musicologists alternate between a closed listening practice, relying exclusively on aural observations, and a multimodal listening practice, where they interact with various music representations and information sources using different media (e.g. metadata about the recordings and performers, sound visualisations, scores, lyrics and performance videos). The spoken parts of broadcast recordings brought historical/extra-musical clues helping to understand music performance practices. Sound visualisation and computational methods fostered the analysis of specific musical expression patterns. We suggest that software designed for musicologists should facilitate switching between closed and multimodal listening modes, interaction with scores and lyrics, and analysis and annotation of speech and music performance using content-based MIR techniques.

1. INTRODUCTION

The interdisciplinary research area of music information retrieval (MIR) has developed from two needs: managing increasing collections of music material in digital form, and solving fundamental problems related to music analysis and perception [1]. Over the past decade, a wide variety of MIR techniques and tools have been developed using various types of music representations (audio, symbolic, visual and metadata). However, as Cunningham [2] points out, they have often been designed based on anecdotal evidence, intuitive feelings, or a priori assumptions of user behaviours and needs. Without a rich understanding of the latter, systems designed using MIR research might not be ideally suited to their potential users. Bridging the gap between the research laboratory and real-world situations is one of the goals of this study. This requires working with specific user groups in order to better understand their activity and how they interact with technologies.

We focused on eliciting some of the strategies used by musicologists to explore musical documents, and the interactions with music-related technologies during this process. Bonardi [3] proposed interesting solutions to improve musicologists' workstations using MIR technologies by examining their needs when analysing the contemporary catalogue at IRCAM's digital library. He stated that the workstations should allow various representations of music (e.g. graphical, sound, and symbolic), listening to recordings while consulting different musical documents ('active' listening), and reading (e.g. the score) and writing using the same media.

In this study, we sought to obtain evidence to test the validity of such statements, and whether they would be relevant in a different context (setting, different types of musicological studies, and musical repertoires). We conducted an ethnographic study based on the observation of musicologists working with classical music recordings at the British Library in London. The ethnographic method is a *qualitative* approach by which findings are not inferred from statistical tests but from the detailed analysis of the behaviours and actions of the participants across a large number of field observations. The outcomes of our research are twofold: they give insights on how to adapt or improve existing digital music technologies to fit user needs better (e.g. MIR techniques, software features, user interface design), and they can raise ideas for the development of new systems.

The remainder of the paper is structured as follows. Section 2 presents the setting and methods of the ethnographic study. Sections 3, 4, and 5 are devoted to thematic analyses based on the observations. We discuss the findings in Section 6 and give a conclusion in Section 7.

2. SETTING AND METHODS

Via the Edison Fellowship scheme, the British Library (BL) encourages musicological studies devoted to the history of recordings of classical music and music in performance, by creating the conditions for concentrated use of the Library's recordings collection to scholars selected on a yearly basis. The BL's sound archive counts more than 3.5 million published and unpublished recordings of sounds and music including many unique historic items. In an era where the Library develops the access and analysis of digitised recordings, and given the close link between MIR and library science (e.g. representation, classification, metadata), the BL is a good setting for investigating the possible roles of music-related technologies in musicological research. We contacted four Edison Fellows with the collaboration of the BL's music department staff, and obtained their consent to participate in the study. The names used in the fieldnotes presented throughout the article are pseudonyms. The group was formed of two British and two American males (average age 38). Their professional activities included research and teaching positions (PhD in musicology, lecturer in music, singing teacher), as well as performance (pianist, singer). Two of them had received training in science and technology. The musical repertoires they studied were varied: early music (e.g. medieval dance, vocal and consort music), classical and romantic music (e.g. art songs, operas, piano solo pieces), and contemporary music (electronic music).

We chose an ethnographic methodology primarily based on participant observation [4]. One of the advantages of observing the actions of participants performed in a concrete setting is that it gives access to *what people do and how* (behaviours) rather than *what people say* (attitudes), the latter being obtained with other qualitative methods such as survey, questionnaire, or interviews. Furthermore, staying for a relatively long period of time in the environment of the group studied fosters the collection of rich details which would otherwise demand a high degree of self-awareness and a great power of recall for people to report out of the context of the activity. To achieve this level of detail, it is necessary to focus on a small number of participants. This tradeoff of quantity for quality is common in disciplines relying on qualitative methods (e.g. psychology) [4, 5].

The observations were made by focusing (i) on the processes underlying musicological research, and (ii) on the relationships with music-related technologies and their roles during such processes. The observations took place in the music department of the BL where the Fellows had a reserved desk space at their disposal. They were conducted by one ethnographer during repeated visits (twice a week, on average) over a period of three months. The observational data were collected by taking fieldnotes using a notebook and pen. Due to the regulations of the British Library and in order to minimize disturbance to the staff, video/audio recordings were not used. Ancillary sources of information were also used in addition to the observations. Ethnographic interviews [4] occurred during the research in the field in order to shed light on specific tasks and to have a deeper understanding of the scope of the studies of the participant. Some of the participants' own working notes were also employed, with their consent. The collected data were analysed using the approach proposed in [6], which draws from methods developed by sociologists following the grounded theory: coding of the fieldnotes (identifying and naming specific analytic dimensions and categories), and analysis by themes which reflect recurrent or underlying patterns of activity.

3. USE OF RECORDINGS

3.1 Retrieval and metadata

Metadata were used to facilitate the retrieval of recordings in the BL's catalogue (by using details such as the record number, the label, or the conductor's name). Additional metadata were fetched during the listening process (see Subsection 4.2), using various sources of information: the knowledge of the Library's curators, the web, the recordings' carriers, liner notes, or accompanying manuscript documents (e.g. a paper card system that an original collector had kept).

3.2 Format and playback technologies

The recordings already digitized were immediately accessible through the British Library Sound Server as MP3 files. When the recordings were unique or held on fragile formats (e.g. reel to reel tapes), the Fellows were provided with analog copies of the recordings or digitized versions on audio CDs, or less commonly, VHS tapes (PCM). The analog formats included reel to reel tapes, compact audio cassette (K7), as well as long-playing (LP) and 78 rpm discs. The recordings held on a physical support were played using dedicated playback equipment connected to an amplifier. The MP3 files from the Sound Server were played from the desktop computer using Windows Media Player Classic. In some cases, they also listened to and analysed owned commercial recordings with their laptops using iTunes to play recordings, and Sonic Visualiser [1] as a player and analysis tool (see Subsection 4.2). For some of the Fellows, the format was not an issue since they were interested in the content of the recordings and not the carrier itself. In that case, they were not bothered by use of MP3 files rather than uncompressed digital or original analog recordings. On the contrary, digital recordings were preferred because the navigation in recordings was made easier and quicker. However, others preferred to deal with recordings in their original format (*"There is more context when you have the original, the labels, how it was held for instance. With most MP3s you do lose something. I do wonder whether sometimes you're losing the core product."*).

[1] `http://www.sonicvisualiser.org/`

4. LISTENING AND OBSERVING

4.1 Listening practices

The listening process lay at the center of the study of the recordings. The musicologists commonly alternated two distinct but complementary practices of listening. In the first listening practice, the analysis of the recordings was performed exclusively through aural observations. The second listening practice was multimodal and characterised by an interactivity with musical, textual or visual documents enriching or modifying the aural observations.

4.1.1 Closed listening

Closed listening was characterised by a careful and focused listening to the recording without using any other source of information than that provided by the sound: *William put his headphones on to listen to Telemann's Concerto in F for 3 recorders, 2 oboes, 2 violins, and continuo, performed by the Early Music Consort. After starting the recording in the CD player, he sat back in his chair, closed his eyes, and listened carefully to the music. A moment later, I noticed that he was tapping the beat with his foot.* This example shows how the aural experience became a physical one (tapping the beat with the foot) while retrieving information about the timing of the musical piece (tempo). In the closed listening mode, the musicologists drew aural observations involving perceptual and cognitive aspects (a recollection of the score, for instance). Either in parallel or shortly after the listening process, they wrote down their aural observations by hand or using a text editor. Typed notes had the advantage that they could be queried quickly by using keywords such as the name of a composer.

4.1.2 Multimodal listening

A different practice was characterised by the use of various music-related documents (e.g. the biography of a composer, information on the recording) and music representations (e.g. scores, feature visualisations) while listening. This listening practice can be described as an active process [3], since it does not just consist of receiving musical information, but is on the contrary based on a set of multimodal interactions between the listeners and musical documents. The advantages of using multiple modalities were an increased access to meaning, uncovering the context of a recording and the intentions of composers, conductors, or performers, and better understanding of the perception of the music. Multimodal listening was performed by varying the media and technologies used to document the musical recordings.

4.2 Documenting the music recordings

4.2.1 Contextual information

In the multimodal listening practice, the musicologists commonly used web resources to seek several types of information related to the recordings: contextual (finding metadata about a musical piece, for instance), bibliographic (music artists' websites, Wikipedia), as well as visual and iconographic (YouTube videos were sometimes used to uncover visual aspects of performance, Google Images was used to provide pictures of specific musicians). Such resources were also used without listening to the recordings.

4.2.2 Scores and lyrics

The online music sheet database from the International Music Score Library Project [2] was often used to retrieve public domain editions of scores, which are provided as scanned images in PDF format. Some of the musicologists read the score using a printed copy, while others used the electronic format and followed the music with the mouse while listening. When they were available, scores were used both in order to retrieve general information such as the key of a piece, and more detailed information through a close analysis of the notes and expressive notation. Singing while reading the score was sometimes used to find the scale used by the composer (e.g. Lydian mode). Scores acted as a reference against which to test whether the intentions of the composer were respected by performers, as the following notes show: *"Seems really consistent with markings in the score. Beautifully sung - singing the note values and generally the dynamics written by Samuel Coleridge-Taylor."*, *"Is much freer with the interpretation of the score. Interpolates a high note at the end and changes the melodic line at the end of the song."* In the case of vocal music compositions, reading the lyrics while listening also helped to follow the musical structure and to understand the expression, as shown in this note describing the timbre of the performer's tones by reference to the lyrics rather than the pitch: *"quite shrill and shaky on 'A wind comes and let me be', and more mellow on 'said it slow'."*

4.2.3 Sound visualisation, acoustical analyses, and time-stretching

Musicologists with previous background in music technologies (coming either from their education, personal training, or from collaboration with computer scientists) also used software (Sonic Visualiser) to analyse and visualise music recordings. The visualisation of the waveform was helpful to navigate digital recordings by jumping between sections that have different dynamics (e.g. between a spoken part and the start of an orchestral part, for instance). Spectrogram representations were used to analyse the subtleties of expressive effect such as the vibrato: *"If I'm looking at*

[2] `http://www.imslp.org/`

a waveform [the one from a tone's partial] and I can see there is vibrato in the note, I hear it much better". Such acoustical analyses helped to understand the perceptual effects experienced when listening: *"The spectrogram shows you that the real skill to her [Emma Kirkby's] vibrato use is that the note starts with very very minimal vibrato. So your mind is fed a very accurate pitch, before the pitch is then decorated by vibrato. So that's why you hear it as such a pure voice, because she's already told you the information about exactly what the note is before it vibrates, so your brain somehow keeps on that central tuning issue during the vibration [...] Whereas singers that immediately start with vibrato, you can never really tell what they're singing."*

Acoustical measurements were performed from the spectrogram representations (a measure tool is provided in Sonic Visualiser) in order to characterise the properties of vibrato (frequency, and pitch extent). These measurements were conducted in a systematic way for various performers by comparing long sustained notes. The resulting quantitative data gave clues to understand or nuance aural observations made on vibrato by other musicologists: *"The minimal vibrato sounds that Munrow listed [...] were all faster and shallower than the other examples of vibrato. When this is combined with Munrow's own explicit disapproval of* constant *vibrato, we begin to understand that he is suggesting a preference for 'controlled' vibrato"*.

The time-stretching technique provided by Sonic Visualiser which preserves the original pitch and timbre was also used to produce slowed-down versions of notes or musical passages. These slowed-down excerpts were played while visualising scrolling spectrogram representations giving the time to the ear and the eye to uncover fine details: *"I knew something was up through listening but I couldn't tell what was up, and then when I visualised ... when I slowed down, more of it made sense, I realised the vibrato was not consistent, but I couldn't work out that it started without vibrato without the spectrogram"*. Spectrogram analyses and time-stretching were also used to validate intuitions obtained with aural observations to explain the technique and expression of a pianist. By looking at the alignment of the notes on the spectrogram while listening to a slowed-down passage, subtle differences of timing between chord notes played by the left and the right hands were noticed.

5. MUSICAL FEATURES

5.1 Instrumentation and tuning

Details about the instrumentation were retrieved in several ways: from the recordings' metadata, from the announcer in the case of broadcast music programmes, or by ear when listening to a musical piece. The choice of instrumentation was an important aspect in the study of historically informed performances of early music (e.g. choice of epoch instruments rather than modern ones), especially since early music scores do not indicate instruments. The recognition of modern versus epoch instruments in musical performances was not a trivial process to perform aurally (*"I'm assuming these are modern instruments, 440 etc."*). Similarly, isolating a specific instrument amongst an ensemble (e.g. the violin in a string ensemble: *"'Quan je voy le duc' - most attractive instrumental piece of collection but horrid scratchy string playing, fiddle?"*), or retrieving the number of musicians playing a part in a specific register (*"Two sopranos?"*) were not easy tasks. The tuning of the instruments was also used to judge musical interpretations (e.g. *"Lamento della Nymfa particularly telling with too many harpsichords I feel - each one slightly out of tune."*, *"Tuning of violins not great in second track"*).

5.2 Musical expression

Various musical features correlated to musical expression were recurrently analysed, including: dynamics (e.g. *"Deller using great sweeping phrases with many dynamic nuances."*), timing (e.g. tapping the beat while listening to increase the sensation of the tempo, measuring the duration of a performance, detailed analyses of pianists' hand asynchronies using spectrograms), timbre (e.g. *"When she sings softer, she doesn't have the same quality. The notes sound mellower."*), pitch (e.g. *"The King's Singers' style hasn't changed much but alto sound is flat!"*), vocal style (e.g. *"Overabundance of rolled 'R's - stylistically ok, but a bit obtrusive in an otherwise beautiful rendition."*), vibrato (e.g. *"Deller consort still has a lot of vibrato in tenor(s) but very good ensemble singing."*), and phrasing (e.g. *"Reminiscent of baroque phrasing rather than renaissance."*).

Musical expression was analysed either by considering a specific performer (e.g. the singer Deller), or by considering an ensemble (e.g. the King's Singers, the tenors), by focusing on the notes, or on phrases, the latter showing the use of different time scales in the analyses. Some features were more difficult to describe solely based on aural observations than others. If dynamics variations seemed to be easily perceived, some variations of pitch and timbre were more difficult to detect confidently (e.g. *"Not sure it stays in tune too well, sinking over the whole perf, less than a semi."*, *"May be because of the choice of quality for notes of the same pitch on two different pieces, the voice doesn't sound the same: it doesn't sound as shrill as it did on the G. May be due to the key D♭."*). Often the expertise of the musicologists as performers was employed to find causal explanations of sound effects based on instrumental techniques: singers were able to associate vocal timbre variations with the vocal technique used to produce them (*"Deller seems to use chest voice for the second, lower, 'Zion'."*), pianists were able to detect timing effects between notes by focusing on the hand technique (hand asynchrony in chords) characteristic of the style of the performer. Musical expression was also described in a critical way by using aesthetic judgments

(*"Soprano sound is rather lovely it must be said"*, *"'desolata' is quite seasick"*, *"I love the bottom of her voice"*, *"Beautiful - very clear rendition"*, *"Very rousing and exhilarating rendition by Webster Booth"*).

6. DISCUSSION

6.1 Visualisation and computational analysis enrich the empirical evidence

Even though, as educated and expert listeners, musicologists were able to perceive extremely fine details, visualisation and computational analysis conveyed empirical evidence which helped them to confirm and prove aural observations (*"The tools on one hand, I don't need them, I could describe that, on the other hand I can't prove it. This tool [Sonic Visualiser] is allowing me to express that in some way it [the finding] is objective."*). As put forward by Cook [7], computational methods bring the potential for musicology to be pursued as a more data-rich discipline. The observations reported in Subsection 4.2.3 show the utility of multiple sources of information to analyse music performance practices. Visualisations and quantitative data retrieved through signal measurements were helpful in discussing, interpreting, or proving hypotheses about qualitative data collected through aural observations. Furthermore, these analyses enabled systematic comparison of the musical expression of various performers in different musical pieces (e.g. measurement of the rate and extent of the vibrato on long sustained notes based on spectrogram analyses) and led to explanations of expressive techniques which could not be reached through aural observations alone (*"You can only hear the pitch aspect of the vibrato as an educated listener with no software or technology."*).

6.2 Cross-modal effects exist between auditory and visual feedback

Many of the examples given in Subsection 4.2.3 also show that the visualisation and listening processes (either at the original speed or using slowed playback) affect each other. For example, the spectrogram helps to hear vibrato much better, the slowed playback of a tone helps to uncover that the vibrato is not constant, while the spectrogram aids in understanding that the variation comes from the fact that the note starts without vibrato. Hence, new empirical evidence emerges from the cross-modal effects between auditory and visual feedback. Visualisation was described by one of the Fellows as a "learning process" (*"Now I've seen the spectrogram, I can only hear it [the vibrato], it's there now ... in my understanding.*"). However, cross-modal effects between auditory and visual feedback also raise a paradox: if visualisation brings to the aural experience an *"increased emphasis on what you can see"*, it concomitantly *"deemphasises what you can't see"*. Therefore the ear may discard relevant aspects when the eye focuses on a spectrogram representation while listening. After performing analyses based on spectrograms, one Fellow noted *"I completely forgot about the bassoon, it feels like it is unimportant now, but I was once struck by it."*. For this reason, being able to listen to a musical piece at first without visuals was deemed to be important, otherwise visualisation may *"irreversibly edit stuff out of your brain that you can't see"*. The designers and users of music feature visualisation software need to be aware of cross-modal interactions which might affect the objectivity of their observations [8].

6.3 Software for musicologists should support closed and multimodal listening practices

We suggest that software designed for assisting musicologists in their analyses of recordings should be in line with their listening practices by supporting both closed and multimodal listening. Due to the cross-modal effects mentioned in the previous section, it would be helpful for the user interface first to provide a closed listening mode without visuals, and then offer the possibility of switching to a more advanced listening mode offering multimodal feedback. The multimodal mode should link the music documents and representations using aural, visual, textual, and symbolic information (see Subsection 4.2). Different software or user interfaces may be needed to handle primary (e.g. scores and sound visualisations) and secondary (e.g. music biographies) information sources.

One way of providing textual and visual information related to a recording (e.g. metadata, pictures) is via semantic web technologies. Linked data offer promising ways to facilitate the retrieval of metadata describing the recordings (date, album art covers, etc.) and the musicians (biographies, photos, etc.). In addition to visualisations of acoustic parameters (see Subsections 4.2.3 and 6.1), the visualisation of scores and/or lyrics within the software would facilitate the analysis of music recordings. Semantic web technologies may also provide ways to retrieve scores from online databases directly from the audio player. Scores could then be used as a reference to compute the performers' expressive deviations using content-based MIR techniques. The visualisation of expressive deviations could help musicologists to determine the extent to which expressive markings in the score are followed in the performance (see the note mentioned earlier: *"Seems really consistent with markings in the score."*), and to characterise the artistic intentions of the conductor and/or performers.

Based on the observations reported in Subsection 4.2, the alignment of scores, lyrics or other time-based metadata to audio recordings could also aid performance practice analysis, by facilitating multimodal listening and providing better navigation of audio documents. For annotation of recordings, the inclusion of text editing functionality into analysis and playback software would be a wel-

come feature, since musicologists generally write down observations while listening. This could indeed be a means to connect notes up with the actual point-in-time of the music which would ease further proof-reading or enrichment of the notes. Controlling the audio playback, with either the keyboard or with a transcription foot pedal, would facilitate tasks such as the transcription of interviews from broadcast recordings including speech and music, and avoid the constant switches between various computer software or different devices which are time-consuming (*"It's so irritating transcribing from a computer file because you're also trying to write on the same computer, so you have to keep going into that program to move the recording back a bit, go back to the word program to type up that sentence more accurately. So [...] if it's my file on my iPod, I can start and stop using a different device than the computer, or here I'm using the CD player."*).

6.4 Can content-based MIR aid musicological study?

Several areas of content-based MIR are relevant for musicological purposes. For instance, automatic speech/music segmentation would help the navigation between spoken and music parts of documentaries and other broadcast material. Speech recognition software would also be of considerable help to automatically transcribe interviews, enabling search of the non-music audio segments for conversations about specific topics or musicians. Regarding the analysis of performance practices, automatic source separation techniques could facilitate separate analysis of the musical expression of different performers or groups of performers (see Section 5.2). Variations of timbre are more difficult to qualify aurally than other variations such as in timing. Therefore MIR techniques improving timbre characterisation (e.g. at the note level) and identification of instrumentation or performers could help answer questions like: *"Is that Janita using some vibrato in the solos?"*

7. CONCLUSION

In this paper, we presented and analysed ethnographic observations of musicologists studying classical music recordings. The observed patterns revealed the importance of: (i) the alternation of closed and multimodal listening modes; (ii) the use of visualisation and computational methods to provide empirical evidence about listeners' impressions; (iii) scores and lyrics acting as a reference in performance analysis; and (iv) web sites and speech recordings supplying historical and extra-musical information.

These findings give clues regarding how to improve software designed for musicologists. Such software should both support closed and multimodal listening, minimising distractions and allowing the user to decide on the display of any feature visualisations during listening. The features of interest for computer-assisted musicology are those characterising artistic choices such as performers' expressive intentions (e.g. tuning, temperament, timing, pitch, timbre, dynamics, articulation and vibrato), most usefully displayed in conjunction with scores and lyrics. Content-based metadata sonification should be handled to facilitate the interpretation of the features (e.g. pitch). Interfaces managing the retrieval of contextual information (e.g. metadata, biographies, articles, pictures) during multimodal listening would benefit the historical approach to musicology. Linked data offers a promising way to connect such extra-musical information with the recordings by exploiting web resources such as the open music encyclopedia MusicBrainz [3] .

8. ACKNOWLEDGMENTS

The authors wish to thank the Edison Fellows and the British Library for their kind participation and help during this study. This study was conducted as part of the RCUK Digital Economy project EP/I001832/1, *Musicology for the Masses* [4] .

9. REFERENCES

[1] J. Futrelle, and J. Stephen Downie: "Interdisciplinary Communities and Research Issues in Music Information Retrieval", *Proceedings of the International Symposium on Music Information Retrieval*, 2001.

[2] S.-J. Cunningham, N. Reeves, and M. Britland: "An Ethnographic Study of Music Information Seeking: Implications for the Design of a Music Digital Library", *Proceedings of the 2003 Joint Conference on Digital Libraries (JCDL'03)*, 2003.

[3] A. Bonardi: "IR for Contemporary Music: What the Musicologists Needs", *Proceedings of the International Symposium on Music Information Retrieval*, 2000.

[4] G. Gobo: *"Doing Ethnography"*, Sage, London, 2008.

[5] R. Collins: *"Theoretical Sociology"*, Harcourt, San Diego, 1988.

[6] R. M. Emerson, R. I. Fretz, and L. L. Shaw: *"Writing Ethnographic Fieldnotes"*, The University of Chicago Press, Chicago, 1995.

[7] N. Cook: "Computational and Comparative Musicology", in *Empirical Musicology: Aims, Methods, Prospects*, Oxford University Press, New York, 2004.

[8] S. Dixon, W. Goebl and E. Cambouropoulos: "Perceptual Smoothness of Tempo in Expressively Performed Music", *Music Perception*, Vol. 23, No. 3, pp. 195–214, 2006.

[3] `http://musicbrainz.org/` and `http://linkedbrainz.c4dmpresents.org/content/linkedbrainz-summary`

[4] `http://www.elec.qmul.ac.uk/digitalmusic/m4m/`