TOWARDS AUTOMATED ANNOTATION OF ACOUSMATIC MUSIC

Volkmar Klien[1,2], Thomas Grill[1,2], and Arthur Flexer[1]

[1]Austrian Research Institute for Artificial Intelligence, Freyung 6/6, A-1010 Vienna, Austria
[2]University for Music and Performing Arts Vienna, Anton v. Webern Platz 1, A-1030 Vienna, Austria

ABSTRACT

At the Austrian Research Institute for Artificial Intelligence (OFAI) we are currently undertaking a two year research project entitled *Towards Automatic Annotation of Electroacoustic Music* [1] investigating the possibilities and potential obstacles for finding (partial) solutions to problems related to computer assisted annotation of electroacoustic music.

Setting aside technological issues pertaining to the relevant fields of signal processing and music information retrieval the paper at hand aims at outlining the reasons behind our choice of *Smalley*'s theory of *spectromorphology* (SM) as our conceptual background, issues pertaining to the role of the annotated score, the formalisation of spectromorphology for automation as well as potential limitations. Given that neither the manual annotation of acousmatic music nor the technical implementation thereof can be seen as straightforward matters, research in this area is still at a very basic level making fully automatic and even fully functional semi-automatic annotation of electroacoustic sound a long-term research goal.

1 INTRODUCTION

In our attempt to create concepts and toolsets for the automated annotation of *acousmatic music*, we use Smalley's theory of *spectromorphology* (SM) [21] as a starting point.

Following Smalley's definition we refer to electronic (or *electroacoustic*) music as *acousmatic* if it is not traditionally note based, does not rely heavily on the listeners' understanding of anecdotal content or their ability to recognize the sounds' physical origins. Even though this holds true for a fair number of 'tape' compositions in a musical tradition often referred to as 'acousmatic', our use of the term includes organized sound from loudspeakers attributed to all kinds of traditions and practices in the sonic arts.

[1] Funded by the *FWF Austrian Science Fund*, Austria's central funding organization for basic research (Project Number: P21247). With additional funding provided by the University for Music and Performing Arts Vienna (*Univision:2* program)

EMS2010, June 21-24, Shanghai, China

Contrary to notated instrumental music, in acousmatic music there exists no pre-segmentation of the heard into units presumed relevant, like notes found in a traditional score. Nevertheless listeners of acousmatic music are usually able to identify musical qualities that are *carriers of meaning* without special preparation, much like in traditional music.

A number of approaches have been made to categorize and describe the origin and nature of these carriers of meaning and find appropriate notions. The theory of *spectromorphology*, building on the basis of *Pierre Schaeffer*'s works, is one of the most influential. It is the theory of temporal unfolding and development of sound spectra and it provides "tools for describing and analyzing listening experience" [21].

It is spectromorphology's embodied approach to musical gesture and its conceptual flexibility in the description of structural levels and dynamic attribution of function that allowed it to resonate with numerous practitioners and theoreticians in the sonic arts. This dynamics – as will be discussed in the paper at hand – can at the same time be seen as the central difficulty in any attempt to systematize and formalize SM.

2 ANNOTATION

2.1 The analysis of acousmatic music

Setting out in the analysis of an individual acousmatic work it cannot be taken for granted what the adequate, most relevant analytical parameters should be. Hence one of the main analytical objectives is to unearth those aspects central to the organization of sound in a given composition. Whereas in instrumental, scored music the analyst more often than not can confidently focus on pitch and rhythm, in electroacoustic composition the structurally most salient matter may well be spatial aspects, gestural development, energy trajectories over time, and others.

Throughout the centuries musicology has developed a rich tool-set of methods for the analysis of instrumental and vocal music bound to the existence of a notated score. Traditional approaches to music theory hence leave us rather helpless in any attempt to analytically describe electronic sound that goes beyond simply modeling traditional musical instruments. The reason for this lies – amongst others – in the fact that acousmatic music does not restrict itself to sonic material that, by convention, has been defined as musical. There are no predefined building blocks or basic musical objects (represented by notes) and no pre-defined or obvious syntax guiding the arrangement of these musical units along time and frequency grids. Given the high output of electronic music, it is surprising that the emphasis of contemporary musical analysis still is on the various genres of instrumental rather than electronic music. In recent years a number of publications (e.g. [20, 17]) presented collections of analyses of electroacoustic compositions. Although practically all of them were created using some form of computational representation of sound (see [1] for a review), the vast majority relied on purely manual annotation of a composition's sound. A number of tools for computer-aided annotation has been developed: *Clam Music Annotator* [2] (Universitat Pompeu Fabra), *ASAnnotation* [3] (IRCAM), *Acousmographe* [4] (INA-GRM), *Sonic Visualiser* [5] (Centre for Digital Music at Queen Mary, University of London) and *iAnalyse* [6] by Pierre Couprie. These are software packages allowing for manual annotation of electroacoustic sound with different levels of support by integrated digital signal processing tools (e.g. for transient or pitch detection).

2.2 The analytical score

It would be difficult to overestimate the role traditional scores played and continue to play in the development of Western art music composition as well as music theory, where it not uncommonly has been viewed as the 'true location' of the musical work (cf. [5, 11], see also [6]). Without going into details of this discussion it will suffice to say that the role of the annotation score will be a fundamentally different one (cf. [8]). The aim of our research for facilitating automated annotation of acousmatic sound (and hence the production of 'scores') is not to finally mend a perceived shortcoming but to detect and describe perceptually relevant musical materials, their interrelation and evolution in time.

The classical musical score, although based on analysis of pitch and metric rhythm is also shaped distinctly by the necessities of instrumental sound production and vice versa. In electroacoustic music with its multitude of production techniques production scores are no longer an imperative and (in practice) hardly ever exist in monolithic form. Hence the annotated listening score in electroacoustic music is completely de-coupled from sound production.

[2] http://clam-project.org/wiki/Music_Annotator

[3] http://recherche.ircam.fr/equipes/analyse-synthese/ASAnnotation/

[4] http://www.ina.fr/entreprise/activites/recherches-musicales/acousmographe.html

[5] http://www.sonicvisualiser.org/

[6] http://www.macmusic.org/software/version.php/lang/en/id/10297/

2.3 Manual annotation and potential benefits of automating annotation

Manual annotation of acousmatic music is extremely time consuming [12], a fact that has prevented broader application and recognition of already existing theoretical frameworks. As has been widely discussed in musicology it would be plainly wrong to presume the existence of one single 'most-correct' analysis (and annotation always is the result of analytical decisions) of a given piece, which to closer approximate we intend to apply computational methods. Automation simply automates things; it does not make them 'more objective'.

Any attempt to formalize and automate a task necessarily puts it and concepts relating to it under added scrutiny. In the case of SM this means to test its toolset's ability to provide clear and unambiguous descriptions of sound independent of personal communication with its added channels of bodily gesture and vocal mimickry of sonic behaviour.

Automated annotation will provide the musicologist with an un-emphatic view of the sonic material to measure his or her own listening experience against and vice versa. This process in itself can provide insight in the workings of the annotation algorithms, the analysed composition as well as the analyst's own listening behaviour. We envisage automated or semi-automated annotation to break new ground in musical analysis by significantly accelerating the process of annotation as well as stabilising the analysis' parameters and results.

Even though individual analyses might legitimately follow rather different strategies, automated annotation will allow for an underlying accumulative process of collecting data on the musical works analysed. Hence automated annotation of acousmatic music will help making acousmatic music research a more data-rich endeavour, which in itself has to be seen as a desideratum (cf.[4]).

2.4 Artistic Practice and Automated Annotation

Analysis, not only in its choices of conceptual tools, but in its individual reading of music is a creative act in itself and as such has always played a role in musicians' individual approaches to music. Automated annotation of electronic sound we envision to constitute a step further towards enabling analysis to take on new roles in the creative process of electronic music making.

Frisius sketches out the implications machine-aided notation of the listening score will have on the role reception plays in musical production. "Music of all kinds, in the context of its listening experience, will be described no more in its abstract visual score, but in its concrete sounding image. All audible hence turns into potential objects of musical analysis. [. . .] The relationship between musical reception and production will change fundamentally, as soon as music has become analyzable to a point enabling the analyst not only to describe, but to experimentally alter it." [7] [7]

2.5 Manual Annotation in the context of Music Information Retrieval (MIR)

In MIR research manual annotation of the audio signal is of crucial importance for the development of algorithms allowing computational systems to connect the purely technical representations of the audio signal to first person descriptions thereof, its human intentionality (cf. [16]). Research into acousmatic music can draw on ample experience concerning methodology and practicalities of manual annotation even though most of MIR research primarily concerns itself with more mainstream forms of music.

While our research into methods for automating annotation of acousmatic music finds a multitude of MIR methods to build on, manual annotation of acousmatic music presents rather specific challenges. Firstly, there is no existing corpus of annotated compositions to draw from. Secondly there exist limitations inherent to manual annotation of acousmatic sound in terms of accuracy in regards to time, pitch and timbre. In dense musical passages (e.g. various spectromorphologies overlapping, quickly moving 'clouds' of short sonic events) exact manual annotation in time becomes a sheer impossibility. While it might be easily possible to aurally isolate a quiet click stream 'behind' the musical 'foreground' it can be completely impossible to annotate these events and check the correctness of these annotations in time, even for expert annotators.

This systemic lack of reliable and exact annotation data presents a fundamental difficulty for any MIR algorithm development. In our search for reliable testing grounds we resorted to using *Chowning*'s composition *Turenas* as one of our first objects of interest. This not only because of its relative sonic clarity and homogeneity, but - the composition being a product of pure synthesis - there exists a production score of it (in Music IV score file format), which was made available to us by John Chowning. This does not mean that our methods intended would depend on production scores to function, but it does provide anecdotal evidence of what convenient things scores really are for musicology. Another helpful aspect about *Turenas* lies in the fact that there exist several published analyses and annotations thereof. ([23, 19, 13, 14])

As an example for the problems faced in manual annotation see fig.1. Below the waveform pane each dot represents the onset of a short sound event. These events grow ever denser (to maximum of approx. 40 onsets per second) until they finally form one extended grainy sound. *Pottier*, in his detailed analysis of the piece [19] refers to the sound morphologies in this section as 'grainy lines'. Which in the context of a musical analysis might well be enough, for our goals outlined above though this remains too general a description.

[7] translation: Volkmar Klien

3 THE SYSTEM OF SPECTROMORPHOLOGY?

SM is not a static sound ontology for the description and classification of sound objects, but rather a set of conceptual tools, a vocabulary for describing the listening experience and its dynamics. It does not propose fixed functional, structural levels or static hierarchies but describes dynamic attributions of functions within the context of the listening experience, directed primarily at *intrinsic* relations within the acousmatic work.

Discussing the evolution of SM Smalley writes: My elaborations of 'motion and growth processes', of 'behaviour' and of 'structural functions', were conceived of as relational frameworks for considering the musical context, and I now think that they are best understood as metaphorical mappings which might simultaneously embody intrinsic and extrinsic views. [8] (Smalley 1999, quoted and translated in Weale 2005.[22])

As Smalley acknowledges, 'intrinsic to music' reveals itself as a less well-defined notion as soon as the various modes of perception are recognized as being heavily integrated. Approaching this issue from semiotics rather than an ecological approach to music perception *Atkinson* [2] reaches a similar conclusion.

Although many of SM's terms for describing sonic behaviour have proven to be very helpful in addressing acousmatic music in human dialogue (which is exactly what they set out to do) these terms cannot be interpreted as strict perceptual categories of sonic behavior. Given the intended conceptual overlap between 'neighboring' categories, e.g. of *drifting* and *floating* generating sufficiently unambiguous annotated test and training sets for the development of machine learning algorithms is in need of further research.

In this SM faces the same problem that any linguistic/symbolic representation of musical sound faces, namely to what extent exactly agreement between different listeners can be reached and to what extent this agreement is determined by cultural aspects. This is one of the reasons behind recent sub-symbolic approaches to musical gesture that emerged in the area of embodied music cognition, musical performance and embodied approaches to new interfaces for musical expression (e.g recent work by *Godøy* [9, 10] and *Leman* [15]).

3.1 Problems of Automating Spectromorphological Description of the Listening Experience

The standard machine learning approach towards classification is to divide a set of annotated examples into training and

[8] In this SM displays similarities to ecological approaches to listening in the context of everyday sound [18] as well as music [3].

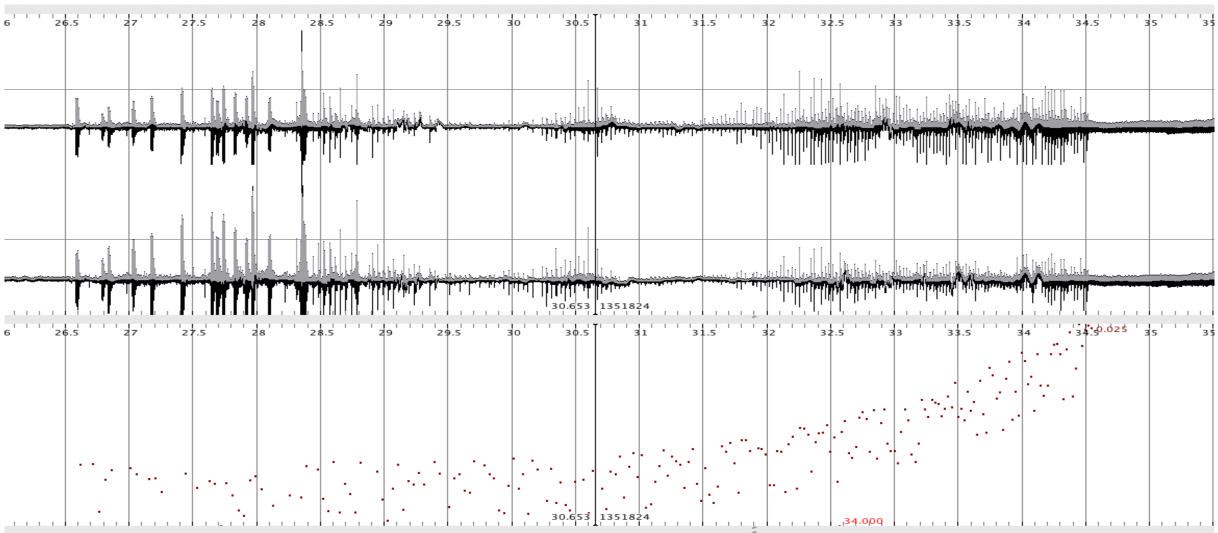

Figure 1. Turenas, seconds 00:26 - 00:35, waveform and event onsets below.

test sets. The training set is used to learn models summing up the relation between the examples and their annotations. The test set is used to evaluate the success in linking examples to their annotations in a fair way.

Even in rather straight-forward musical situations of e.g. tempo estimation or genre classification of popular music the quality of training and test sets is one of the central concerns. It is currently unclear how such sets for SM descriptors can be accumulated, especially given the fact that in real life situations (i.e. acousmatic compositions) hardly any one single spectromorphology appears ‘solo’.

It might be promising to ask several acousmatic composers to produce their individual sonic variants of SM’s classifiers. This would still result in an artificial test set. This test set would need to be seen as a product of its times and would have sounded rather different in the 1970s than it would have in the 1990s. What would emerge to be the signals’ invariants?

4 CONCLUSION

The review of the issues at stake presents a rather complex situation on a conceptual level regarding the taxonomy of SM as well as its linking to the practicalities of automation. Thus we believe that even semi-automatic annotation along the lines of SM needs to be seen as a long-term goal. We are convinced though that even modest results like the implementation of bootstrapping methods for the interactive identification and annotation of groups of similar sonic material (constituting preparatory work to spectromorphological annotation) would be a valuable research contribution.

5 REFERENCES

[1] Norman Adams. *Analytical Methods of Electroacoustic Music*, chapter Visualization of Musical Signals, pages 13–28. Routledge, 2006.

[2] Simon Atkinson. Interpretation and musical signification in acousmatic listening. *Organised Sound*, 12(2):113–122, 2007.

[3] Eric F. Clarke. *Ways of Listening. An Ecological Approach to the Perception of Musical Meaning*. Oxford University Press, 2005.

[4] Eric F. Clarke and Nicholas Cook. *Introduction: What is Empirical Musicology?*, pages 3–14. Oxford University Press, 2004.

[5] Carl Dahlhaus. *Notation neuer Musik*, chapter Notenschrift heute, pages 9–34. Darmstädter Beiträge zur Neuen Musik. Schott, 1965.

[6] F. Delalande. The technological era of ‘sound’: a challenge for musicology and a new range of social practices. *Organised Sound*, 12(3):251–258, 2007.

[7] Rudolf Frisius. *Forum Analyse: Medienspezifische Analyse - Analyse medienspezifischer Musik; Einführung*, volume Konzert-Klangkunst-Computer, Wandel der musikalischen Wirklichkeit., pages 170–171. Schott, 2002.

[8] Évelyne Gayou. Analysing and transcribing electroacoustic music: the experience of the portraits polychromes of grm. *Organised Sound*, 11(2):125–129, 2006.

[9] Rolf Inge Godøy. Gestural-sonorous objects: embodied extensions of schaeffer's conceptual apparatus. *Organised Sound*, 11(2):149–157, 2006.

[10] Rolf Inge Godøy and Marc Leman, editors. *Musical Gestures: Sound, Movement, and Meaning*. Routledge, 2009.

[11] Nelson Goodman. *Languages of Art; An Approach to a Theory of Symbols*. Harvester Press, 1976/1981.

[12] D. Hirst. *The development of a cognitive framework for the analysis of acousmatic music*. PhD thesis, University of Melbourne, 2006.

[13] Luis Jure. Escuchando turenas de john chowning. http://www.eumus.edu.uy/revista/nro1/jure.html, 2004.

[14] Luis Jure. Resintetizando turenas. http://www.eumus.edu.uy/docentes/jure/pubs/resintetizando_turenas/, 2004.

[15] Marc Leman. *Embodied Music Cognition and Mediation Technology*. MIT Press, 2007.

[16] Micheline Lesaffre, Marc Leman, Bernard de Baets, and Jean-Pierre Martens. Methodological considerations concerning manual annotation of musical audio in function of algorithm development. In *Proceedings of the International Conference on Music Information Retrieval*, 2004.

[17] T. Licata. *Electroacoustic Music: Analytical Perspectives*, volume 63 of *Contributions to the Study of Music and Dance*, chapter Introduction, pages xxi – xxiv. Greenwood, 2002.

[18] John G. Neuhoff, editor. *Ecological Psychoacoustics*. Elsevier Academic Press, 2004.

[19] Laurent Pottier. *John Chowning, Portraits polychromes*, chapter Turenas, analyse. Éditions Michel de Maule, 2005.

[20] M. Simoni. *Analytical Methods of Electroacoustic Music*, chapter Acknowledgments, page vii. Routledge, 2006.

[21] Denis Smalley. Spectromorphology: explaining sound-shapes. *Organised Sound*, 2(2):107–26, 1997.

[22] Robert Weale. *The Intention/Reception Project: Investigating the Relationship Between Composer Intention and Listener Response in Electroacoustic Compositions*. PhD thesis, De Montfort University, Leicester, UK, 2005.

[23] Bijan Zelli. *Reale und virtuelle Räume in der Computermusik*. PhD thesis, Technische Universität Berlin, 2001.