
Telling Sounds: Annotating and Connecting Audio(visual) Sources for Musicological Research

Julia Jaklin^{*1}, Elias Berner^{*1}, Peter Provaznik¹, and Matej Santi¹

¹Universität für Musik und darstellende Kunst Wien – Austria

Abstract

The research project *Telling Sounds*^[1] is concerned with the impact of the digital availability of audio(visual) (AV) sources regarding new approaches in the field of musicology. The embedding of music in radio programs, reports, documentaries and films points to the significance of these sources for a music history beyond the grand narrative of musical "heroes" encountered in cultural centers (like Vienna). The sources in question contain different kinds of music, in a variety of contexts: the audio(visual) elements refer to different times, institutions, events, places, persons, repertoires, and sounds respectively, whose multiplicity, diversity, and spread across time and space challenge a merely linear understanding of history. Striving to link such sources across the boundaries of various collections and archives calls for a combination of methods, taken from oral history, media- and film-studies, music analysis, cultural-studies-informed-musicology^[2] and performance studies.

An important objective of our project is developing a "Tool", a research software for capturing and visualizing the interaction of music and its contexts^[3], across various audio and audio-visual sources, across different points in time, and involving certain places and agents. This Tool aims to provide a collaborative means of building a corpus of relevant AV material, as well as enriching existing metadata with annotations about the contents (such as who is talking, or which pieces of music can be heard). Material comes from the collections of our partners, Österreichische Mediathek^[4], Phonogrammarchiv of the Austrian Academy of Sciences^[5], or ORF (the Austrian national broadcasting company), but could just as easily come from YouTube and other online platforms.

AV sources ("Clips") are chosen and categorized based on case studies and their respective research questions. The more contributions there are, the more value can be provided to (music) researchers working with audio(visual) sources, by helping them find relevant material and – facilitated by a graphical representation of the network – revealing previously non-obvious connections while also highlighting the non-linear, often a-temporal and generally networked character of relationships between Clips and the various entities that feature in them. At the same time, searching among the collected entries should be possible, as well as performing queries, such as "Which Clips include a visual of Leonard Bernstein, performing Beethoven's music?", "Which kinds of music are heard in the context of National Socialism?", "Which images of identity are communicated in the wake of the 1970's Beethoven celebrations?", "How is Austro-Pop constructed as a site of memory?" or more generically, "Which entities (and sounds) are appearing together (i.e. at the same time, in the same Clip)?".

*Speaker

From the plethora of the captured entities that appear in a given Clip, more complex interpretations, meanings, and connotations arise, represented in our Tool by the "Topic" concept. Topics may include historical periods, political ideologies, or *lieux de mémoire* (i.e. sites of memory). Since there may be multiple researchers working with the same AV sources and the applicability of a given Topic may often be considered to be subjective or dependent on the context of a given case study, the system needs to support differing opinions by clearly expressing who "said" what. At the same time, it should be possible to reference external bibliographical sources, and auxiliary materials, such as photos or documents.

The implementation of the Tool takes the form of a webapp, whose backend is centered around RDF. We are not importing and storing the AV Clips themselves, but merely pointing to them, capturing only their metadata. Persons, places and other entities are also represented in the system, building on existing data from GND, GeoNames, or possibly WikiData or DBpedia, whenever it is feasible (in order to make data entry less tedious and manual). The connections between Clips and entities, however, have to be entered manually. As noted above, the Tool should not only capture which entities are appearing in a given Clip, but also *in what way* they are appearing. How detailed and expressive these connections need to be, is an area of ongoing experimentation and deliberation. Is it enough to be able to say: "Leonard Bernstein is performing; Beethoven's Op. 15 concerto can be heard.", or do we want to be able to say: "Leonard Bernstein is seen and heard performing Beethoven's Op. 15 concerto together with the Vienna Philharmonic, conducting and playing piano.", or something in between?

The data backend is based around a triplestore. RDF, itself being a graph data model, offers lots of flexibility, is well suited for integration with other systems, and comes with a rich set of existing ontologies and vocabularies. RDF's query language SPARQL provides a lot of expressiveness, allowing things like traversing multiple instances of a property, or querying (thanks to sub- and super-property relationships and inference) for the more general case while still being able to store the specifics of a given connection.

So far, transforming these ideas and requirements into an actual working system has proven to be anything but trivial – in particular, striking a balance between the underlying data model's expressiveness and complexity is a key issue. User experience is also a major consideration, as is, of course, making the resulting data accessible and useful to others. We would like to present the current status of our modeling and implementation efforts and discuss challenges, and are looking forward to comments and suggestions.

¹

: <https://mdw.ac.at/imi/tellingsounds>

²

: cf. Kramer, Lawrence (1990): *Music as Cultural Practice, 1800-1900*. University of California Press., among others.

³

: As already discussed by Nicholas Cook more than twenty years ago: Cook, Nicholas (1998): *Analysing Musical Multimedia*. Oxford University Press.

⁴

: <https://www.mediathek.at>

⁵

: <https://www.oeaw.ac.at/phonogrammarchiv/>

Keywords: music history, media, audiovisual media, digital musicology, semantic web, linked open data, annotations, data modeling, visualization