

TOWARD MUSIC STRUCTURE ANNOTATION

Geoffroy Peeters

Ircam Analysis/ Synthesis Team - CNRS STMS

A large part of present-day “Music Structure” research is devoted to the improvement of algorithms, through the improvement of recognition scores or to the definition of new measures of performances. But a question that should also be asked is “how pertinent is the structure annotation that is used for those evaluations?”. This involves a precise definition of the annotation process and thinking about the annotation’s relevance. In this page, we discuss the various possible definitions of Music structure annotation and introduce a set of concepts which form a “multi-dimensional” description of “Music Structure”.

Music Structure based on “musical role”: One can rely on the choice of assigning labels according to the “musical role” that a part plays in a song (“introduction”, “verse”, “chorus”, “bridge”, “ending”). However, in this case we merge several notions. “Intro” and “Outro” refer to positions in the time axis of the track (sometimes the Intro or Outro is actually the “chorus”). Also they can be several versions of the “verse” and “chorus” (hence some test-sets use labels such as “verse A”, “verse B” which unfortunately mixes “musical role” with “content similarity” notions). The definition of the chorus and of the verse is not always clear (this is the case for Rap or R’n’B music).

Music Structure based on “acoustical similarity”: One can rely on the choice of assigning labels according to the acoustical similarity between parts. Two parts are similar if they are identical (such as in Moby “Natural Blues”), but what about if there is a small variation? a) How to quantify this variation? Is the timbre similarity more important than the harmonic or rhythmic ones? Is an instrument more important than an other? This poses the problem of the point-of-view used to define the acoustical similarity. b) Then how do we go from the quantified variation to the binary decision “they have the same (different) label(s)”. This poses the problem of the choice of a threshold, from which depends the number of different labels used.

Music Structure based on “instrument role”: One can rely on the choice of assigning labels according to the instrumentation of the track. In this, we describe the location of the lead singer parts, the (solo) guitar parts, ... This description however provides few insights into the global structure of the track. Furthermore providing the identity of the instrument will require a huge number of labels (guitar= classical? folk? electric? wha-wha?). In this case it is more useful to describe the “role” plays by this specific instrument in the track, such as Primary Lead (the obvious front-men singer or instrument), Secondary Lead (the backing singer or side-man more generally). We call it “instrument role” in the following.

Music Structure based on perceptual tests: One can also rely on perceptual tests to find the average human perception of the musical structure. Apart from the fact that this approach is very costly, an other problem comes from

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2009 International Society for Music Information Retrieval.

Emmanuel Deruty

Ircam Analysis/ Synthesis Team - CNRS STMS

the fact that, in the “Music Structure” case, the labels used by people to describe a track are usually not shared.

Proposed Music Structure: multi-dimensional representation: The main idea of the proposed description is to use simultaneously (but independently) the various view-points: “acoustical similarity”, “musical role” and “instrument role”. A track is formed: • by a set of Constitutive Solid Loops (CSLoop) which represent a “musical phrase” or a “musical exposition” (a succession of chords). CSLoop with similar ID represent the same “musical phrase” although large variation can occur between them. Two CSLoops with the same ID can follow each other if the “musical phrase” is repeated twice successively. • over which are super-imposed variations of the CSLoops ID. For example the same CSLoop which occurs in a lighter version (for example without the drum or without the bass) is indicated by “-”; if it is in a stronger version (for example with an extra second guitar) by “++”. • over which are superimposed important “instrument roles”: such as presence of the primary leads (lead singer in popular music, lead instrument in jazz or electro music), other leads (choir, other lead instruments or melodic sample) or solo mode (electric-guitar solo, jazz chorus solo, ...) • and which plays a “musical role” (intro, outro, transition, obvious chorus, or solo). The track is therefore decomposed simultaneously on these various view-points. When a part is too complex to be described, it is annotated as ComplexMode. The mandatory decomposition is the CSLoop description. When a CSLoop is an obvious chorus it is annotated as “chorus”. When it is not obvious, it is not annotated as “chorus” but it is still annotated as the repetition of the occurrence of a specific CSLoop, with PrimaryLead and OtherLead (Choir) which are distinctive elements. In order to solve the segment sub-division problem, markers can be placed inside a CSLoop segment to indicate further possible sub-divisions. Two types of markers can be placed (V1 and V2) indicating respectively similarity and dissimilarity between the parts on the left and on the right of the marker. An example from the 300-tracks test-set annotated using this method is represented in Figure 1.

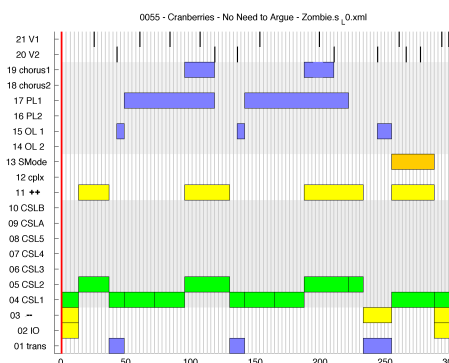


Figure 1. Cranberries “Zombie”.

Acknowledgments: This work was partly supported by “Quaero” Programme, funded by OSEO, French State agency for innovation